

# Automated knowledge base construction

## 7. Commonsense Knowledge (CSK)

Simon Razniewski  
Summer term 2022

# Interactive pattern matching with spaCy

- Interactive tool for exploring pattern matches
  - <https://explosion.ai/demos/matcher>
- General documentation:
  - <https://spacy.io/usage/rule-based-matching>

# Background

- So far largely (implicitly): **Small data**
  - First sentence from an entity description, one Wikipedia article, one Fandom infobox, ...
- More often than not:
  - **Huge set of possible sources**
  - Precision-orientation may suggest filtering, selecting
    - Semistructured content >> Wikipedia >> General web
  - If no premium resources exist/high recall is desired  
→ Leads to the topic of **extraction consolidation**

# Consolidation need

- *bornIn(John, Paris) + bornIn(John, London)*

(single-value conflict)

- *hasParent(Mary, John) + hasParent(John, Mary)*

(antisymmetric relation violated)

- *hasParent(Mary, {John1, John2, ... John377})*

(implausible counts)

- *bornIn/raisedIn/livedIn/diedIn(John, Sydney) + positionHeld(John, Prime Minister of Canada)*

(spatial implausibility)

... temporal implausibility

... topical implausibility

...

# Extraction consolidation

- Family of approaches
  - Tresholding
  - Constraint reasoning
  - Multi-source validation
- Details in
  - Section 8.5 of course textbook (see website)
  - Slides of lecture 8 of old course ([link](#))
- Today: **Application domain** of AKBC where multi-source extraction and consolidation is essential:  
**Commonsense knowledge**

# Outline

## **1. Introduction to CSK**

1. What is CSK?
2. Why is it important?
3. How to represent it?
4. What makes it challenging?

## 2. Crowdsourced CSKB construction

## 3. Text-extraction for **ACSKB** construction

## 4. CSKBs: Summary and Outlook

# What is commonsense knowledge?

Definition 1 (by commonality):

**Knowledge shared by most humans**

- Possible qualifications
  - Across cultures
  - From early in life (=children)
- E.g., elementary school exam questions
  - <http://data.allenai.org/ai2-science-questions>

# What is commonsense knowledge?

Definition 2 (by knowledge type):

**Knowledge about concepts and events**

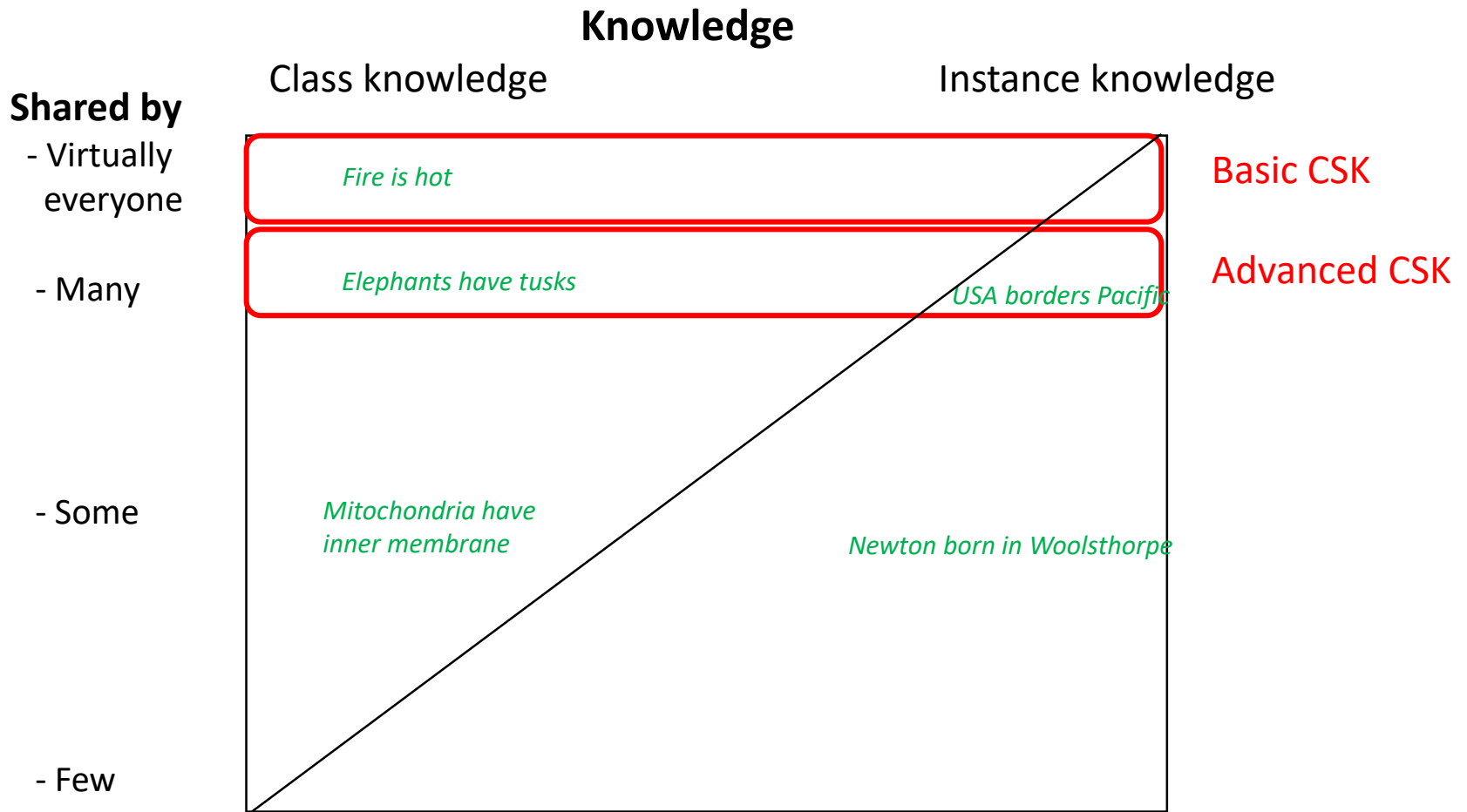
- Concepts: *City, footballer, organization*
- Events: *Football match, birthday party*
- Differentiation from encyclopedic knowledge on instances
  - Instances: *Saarbrücken, Ronaldo, Manchester United*



# Definition Pro/Con

- Definition 1 (by commonality):
  - *Popsicle, is, frozen – only known in North America*
  - *Lion is dangerous/cute - depends whom you ask*
  - Inclusion/exclusion decision challenging
- Definition 2 (by knowledge type):
  - *Apple MacBook, Ford Model T*
  - Class/instance not trivial to separate
  - *USA borders Pacific Ocean – excluded as instance knowledge*
  - *Mitochondria, hasPart, inner membrane – not common knowledge*
  - Open-ended
    - Can be somewhat mitigated by ranking-based evaluation

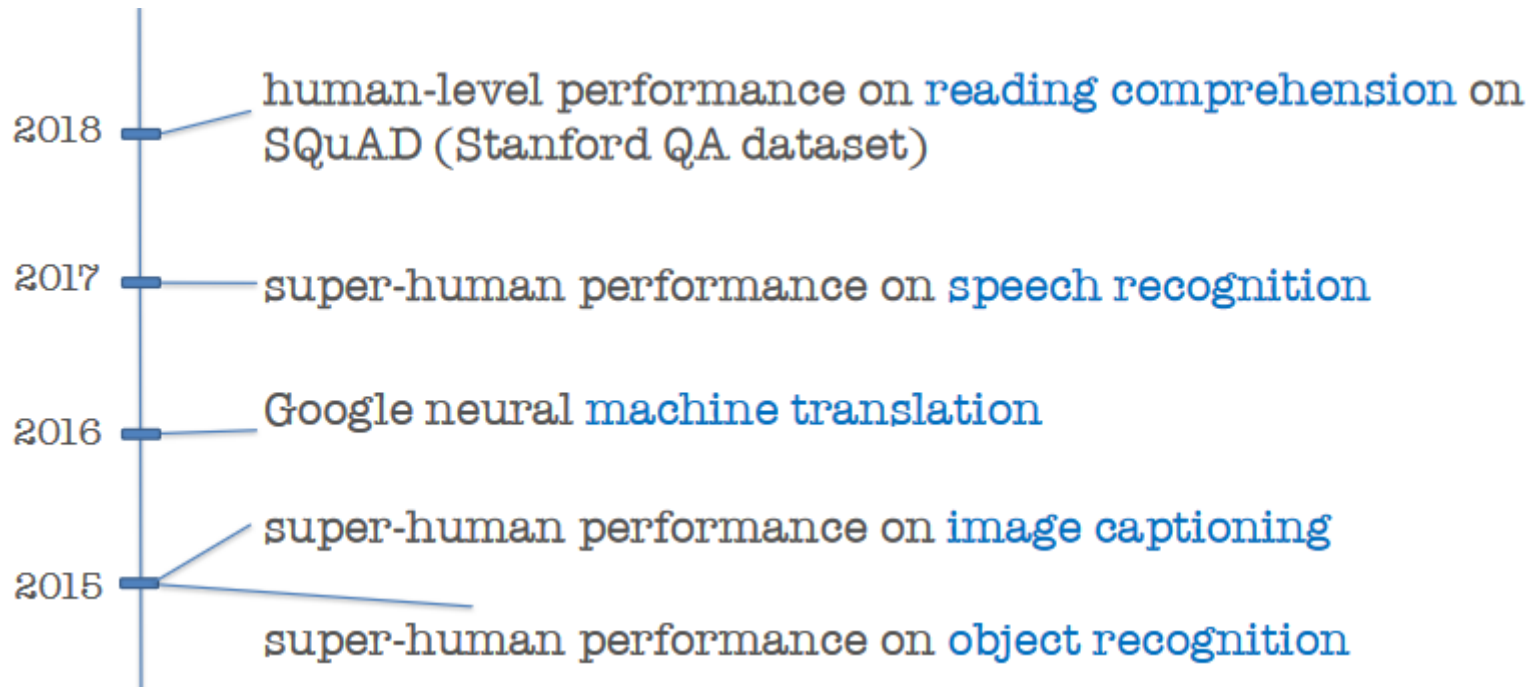
# Definition: Merger



# Examples of CSK

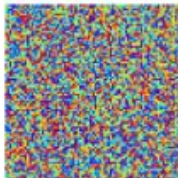
- Taxonomical
  - *Elephant, isA, mammal*
- Properties
  - *Elephant, lives in, Savanna*
- Parts
  - *Elephants, hasPart, trunk*
- Measures
  - *Adult elephant, weight, ~2..5 tons*
  - *Elephant, lifespan, ~60 years*
- Activities
  - *Seeing elephant, requires, go to zoo*
  - *Go to zoo, subevent, buy ticket*
  - *Go to zoo, typicalDuration, 2 hours*

# Why CSK? Amazing progress without





+



=



Giant panda  
Object  
Recognition

Gibbon

Szegedy et al,  
2014....



VQA

Jabri et al,  
2017



A horse standing in the grass.

Captioning

MacLeod  
et al, 2017

How are you  
doing?



I don't know.

Dialogue

Li et al,  
2016



I don't know. I  
don't know. I  
don't know.

Open-ended

Generation

Holtzman  
et al, 2018

.... Nikola Tesla moved to  
Prague in 1880. ... **Tadakatsu**  
**moved to Chicago in 1881.**

Where did Tesla move in  
1880? **Chicago**

QA

Jia et al,  
2017

Solving only a "dataset"  
without solving the underlying "task"!

# Importance of CSK

**Reusable** and **scrutable** asset for a range of AI tasks

- **Reusable:**
  - CSK can be plugged into a range of tasks, e.g., QA, dialogue, object recognition, text generation, ...
  - Contrasts with typical end-to-end learning
- **Scrutable:**
  - Humans can inspect, add and **remove** content
    - Relevant in applications where errors are costly
    - Relevant in applications at risk of bias/discrimination
  - Humans can inspect discrete statements used for reasoning
    - Relevant for debugging complex downstream use cases
  - Contrasts with end-to-end learning and pretrained language models

# Knowledge representation challenges

- Encyclopedic KBs: Typically binary truth notion
  - *Trump, born in, NY*
  - *House of Cards, producer, Netflix*
  - *New York, mayor, Bloomberg, [2002-2013]*
- CSK: Generalizes across subjects
  - *Lions, have, manes* - percentage?
- Fuzzy time notion
  - *Lions, drink, milk* - when?
- Spatial and cultural context
  - *Lion, is, cute*
  - *Elk, usedFor, transport*

# Linguistics - Generics

- It is complicated
  - *Ducks lay eggs >> Ducks are female*
    - Even though former set is a subset of latter
  - *Dinosaurs are extinct/Elephants are biggest land animals*
    - Not applicable to individuals



# Epistemic logics

- *“Zoo visitors believe lions are cute”*
- *“Rural dwellers believe lions are dangerous”*
  
- Used in CycL
  - Reification
  - Modals for belief and desires

[CYC: Towards programs with common sense, Lenat et al., 1990]

# Episodic logics

KNext, Lore projects in early 2000s

```
(REP . RONNIE FLIPPO ( D . , ALA . ) , ONE OF THE MEMBERS  
OF THE DELEGATION , SAYS 0 HE WAS PARTICULARLY IMPRESSED  
*-1 BY MR . KRENZ 'S READY ADMISSION THAT EAST GERMANY  
NEEDED *-2 TO CHANGE .)
```

```
AN ELECTED-REPRESENTATIVE MAY SAY A PROPOSITION.  
A DELEGATION MAY HAVE MEMBERS.  
A MALE-INDIVIDUAL MAY BE IMPRESS -ED BY AN ADMISSION.  
AN ADMISSION CAN BE READY.  
A COUNTRY MAY NEED TO CHANGE.
```

```
((:I (:Q DET ELECTED-REPRESENTATIVE) SAY[V] (:Q DET PROPOS))  
(:I (:Q DET DELEGATION[N]) HAVE[V]  
(:Q DET (:F PLUR MEMBER[N])))  
(:I (:Q DET MALE-INDIVIDUAL) (:F BE[PASV] IMPRESS[V])  
(:P BY[P] (:Q DET ADMISSION[N])))  
(:I (:Q DET ADMISSION[N]) READY[A])  
(:I (:Q DET COUNTRY) NEED[V] (:F KA CHANGE[V])))
```

LORE

BROWSE KNOWLEDGE

QUERY RESULTS: ELEPHANT

7,599 results in the active KB.

Many-or-some elephants pertain to a room.

Many elephants are wild occasionally.

Many-or-some elephants are wild.

Many elephants pertain to a person.

Many people have an elephant.

A herd can be constituted of or filled with elephants.

Many elephants pertain to a male.

Many-or-some elephants are female.

Schubert, Lenhart. "Can we derive general world knowledge from texts." *HLT 2002*

# Graded formalisms

- Heuristic level in CycL

- *True, default true, unknown, default false and false as statement labels*

- Ordinal grades

[Schubert and Tong, NAACL 2003]

1. SEEMS LIKE A REASONABLE GENERAL CLAIM (Of course. Yes.)  
A grand-jury may say a proposition. A report can be favorable.
2. SEEMS REASONABLE BUT EXTREMELY SPECIFIC OR OBSCURE (I suppose so)  
A surgeon may carry a cage. Gladiator pecs can be Reeves-type.
3. SEEMS VACUOUS (That's not saying anything)  
A thing can be a hen. A skiff can be nearest.
4. SEEMS FALSE (No. I don't think so. Hardly)  
A square can be round. Individual -s may have a world.
5. SOMETHING IS OBVIOUSLY MISSING (Give me a complete sentence)  
A person may ask. A male-individual may attach an importance.
6. HARD TO JUDGE (Huh?? How do you mean that? I don't know.)  
A female-individual can be psychic. Supervision can be with a company.

- Simplified

[Zhang et al., TACL 2017]

- *Very likely*
- *Likely*
- *Plausible*
- *Technically possible*
- *Impossible*

Sam bought a new clock $\rightsquigarrow$ <b>The clock runs</b>
Dave found an axe in his garage $\rightsquigarrow$ <b>A car is parked in the garage</b>
Tom was accidentally shot by his teammate in the army $\rightsquigarrow$ <b>The teammate dies</b>
Two friends were in a heated game of checkers $\rightsquigarrow$ <b>A person shoots the checkers</b>
My friends and I decided to go swimming in the ocean $\rightsquigarrow$ <b>The ocean is carbonated</b>

# Graded formalisms (2)

- Dice [Chalier et al., AKBC 2020]
  - 4 dimensions

	Plausible	Typical	Salient	Remarkable
Lions; eat; chicken	✓			
Lions; attack; humans	✓		✓	✓
Lions; drink; water	✓	✓		

# KR - state of the art

- Expressive proposals exist
  - Modal, epistemic, episodic logic
- Instantiation hard
  - Sparse realization in natural language
  - Correct extraction nontrivial
- Most projects:  
Pragmatic choice of (subject, predicate, object) triples with a single score

Lion, hunts, zebra – 0.73

Lion, drinks, milk – 0.45

# Triples and done?

- Still major design decisions left!

1. Fixed or open set of predicates
2. Subject range
3. Object range

- Fixed vs. open predicates

- E.g., ConceptNet: ~25 predicates (isCapableOf, requires, isA) vs. TupleKB ~1000 textual phrases

- Subjects: Strings or disambiguated terms?

- Lynx vs. lynx vs. lynx



- Granularity and modifiers

- Elephant, Foraging elephant? Newborn elephant?

- Objects: Entities or open phrases?

- Politician, isCapableOf, promise that impossible things will happen

# Outline

1. Introduction to CSK
- 2. Crowdsourced CSKB construction**
3. Text-extraction for ACSKB construction
4. CSKBs: Summary and Outlook

# ~~Crowdsourcing~~ Expert annotation

- **WordNet** [Miller and Fellbaum, ~80s]
  - Lexical resource still popular today
  - Is-A, synonym, partOf
  - ~200k word senses
  - 2 expert annotators (=authors)
  - Limitations: Imbalance, idiosyncrasies, lack of scores/ranking
- **Cyc** [Lenat, ~80s]
  - CSK, world knowledge, rules
  - Hired experts on specific domains
  - >1000 person-years of effort estimated

- Next projects: Harness power of laypeople

WordNet Search - 3.1  
- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations  
Display options for sense: (gloss) "an example sentence"

**Noun**

- **S: (n) elephant** (five-toed pachyderm)
  - [direct hyponym / full hyponym](#)
    - **S: (n) rogue elephant** (a wild and vicious elephant separated from the herd)
    - **S: (n) Indian elephant, *Elephas maximus*** (Asian elephant having smaller ears and tusks primarily in the male)
      - **S: (n) white elephant** (albinic Indian elephant; rare and sometimes venerated in east Asia)
    - **S: (n) African elephant, *Loxodonta africana*** (an elephant native to Africa having enormous flapping ears and ivory tusks)
    - **S: (n) mammoth** (any of numerous extinct elephants widely distributed in the Pleistocene; extremely large with hairy coats and long upcurved tusks)
      - **S: (n) woolly mammoth, northern mammoth, *Mammuthus primigenius*** (very hairy mammoth common in colder portions of the northern hemisphere)



# Open Mind Common Sense / ConceptNet

- Havasi et al., MIT ~1999 [<http://conceptnet.io/>]
- CSK for ~25 relations
  - Construction statistics
    - ~14k volunteers filled in sentences with blanks
    - ~700 000 English sentences
    - NLP tools: 300 000 concepts and 1.6 million assertions

elephant is a type of...	Location of elephant	elephant has...	elephant is capable of...
<a href="#">en</a> a pachiderm →	<a href="#">en</a> a circus →	<a href="#">en</a> a trunk →	<a href="#">en</a> carry a trunk →
<a href="#">en</a> pachyderm (n, animal) →	<a href="#">en</a> Africa →	<a href="#">en</a> tusks →	<a href="#">en</a> forget to go on the paper →
<a href="#">en</a> proboscidean (n, animal) →	<a href="#">en</a> a zoo →	<a href="#">en</a> four legs →	<a href="#">en</a> lift logs from the ground →
<a href="#">en</a> emblem (n, communication) →	<a href="#">en</a> India →	<a href="#">en</a> a long nose called "trunk" →	<a href="#">en</a> to lift the tree →
<a href="#">en</a> pack animal (n) →		<a href="#">en</a> a nose →	<a href="#">en</a> remember water sources →
<a href="#">en</a> animal →			<a href="#">en</a> visit the grocery store →
<a href="#">en</a> elephantidae (n) →			<a href="#">en</a> weigh up to 14000 pounds →
<a href="#">en</a> herd animal (n) →			<a href="#">en</a> weight 1000 kilos →

[ConceptNet — a practical commonsense reasoning tool-kit, Liu and Singh, 2004]

# Open Mind Common Sense / ConceptNet

## Knowledge about ocean

Similar objects to **ocean**: sea, water, beaches, aquarium, lake

### An inquiring mind wants to know...

Is on the ocean somewhere that people can be?  
Yes / No / Doesn't make sense / Why do you ask?

You would find  near the ocean.  
Teach OpenMind

Is on the ocean somewhere that coral reefs can be?  
Yes / No / Doesn't make sense / Why do you ask?

ocean is a kind of .  
Teach OpenMind

Would you find an ocean in a pool?  
Yes / No / Doesn't make sense / Why do you ask?

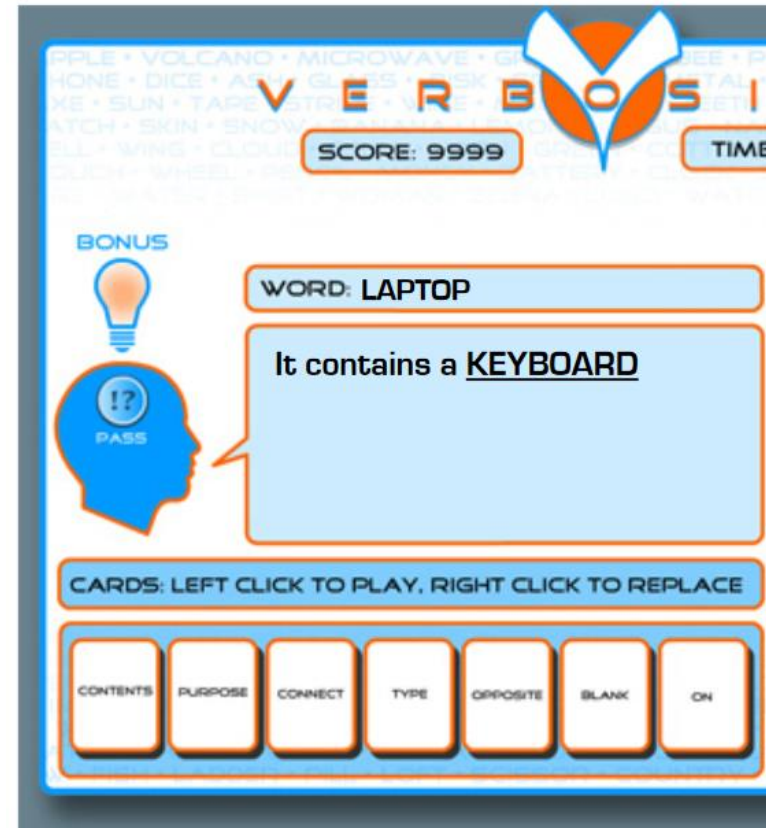
an ocean is used for .  
Teach OpenMind

Is on the ocean somewhere that seagulls can be?  
Yes / No / Doesn't make sense / Why do you ask?

ocean can be .  
Teach OpenMind

# Verbosity

- 2-player game inspired by Taboo
- Narrator must describe a word by filling blanks in templates
  - is a kind of     . *Allows for hierarchical categorization.*
  - is used for     . *Provides information about the purpose of a word.*
  - is typically near/in/on      (three templates). *Provide spatial data.*
  - is the opposite of      /      is related to      (two templates). *Provide data about basic relations between words.*
- Templates give rise to CSK assertions
- Verification via automated narrator that replays human assertions
- Used to feed ConceptNet
- High-quality: 85% of sentences rated as correct by 6/6 annotators

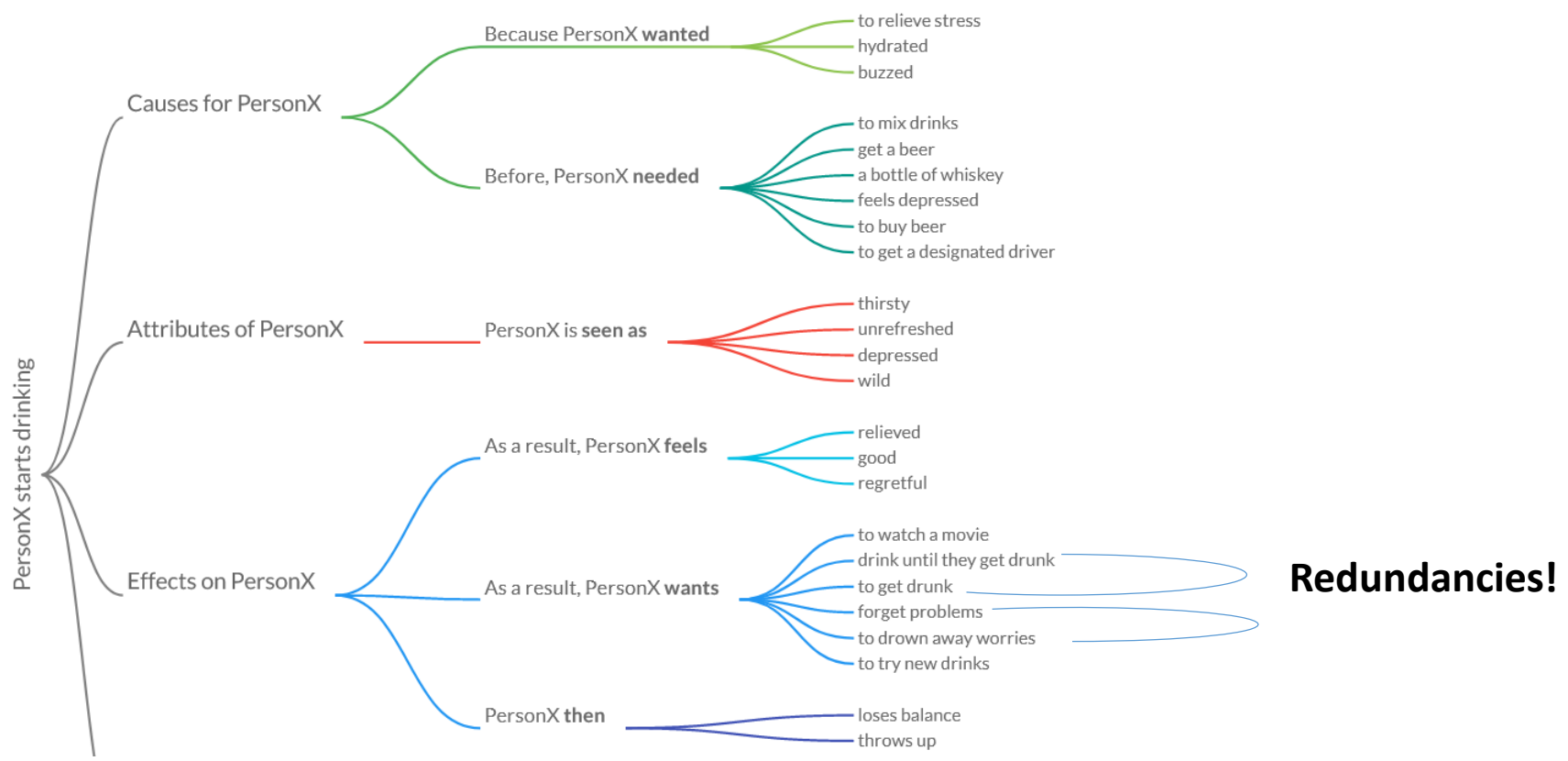


[Ahn et al., 2006]

# Atomic

- Targets event knowledge

[https://mosaickg.apps.allenai.org/kg\\_atomic](https://mosaickg.apps.allenai.org/kg_atomic)



# Atomic

- Archetype of large-scale paid crowdsourcing
- Subjects from text extraction (24k event phrases)
- Statement creation
  - 3 workers per subject
  - Free-form interface
  - ~12\$/hour
  - 300k statements
  - 3\*5 minute/subject (?) → ~\$100k cost

**Event**

PersonX pays PersonY a compliment

## Before

1. Does PersonX typically **need** to do anything **before** this event?

## After

2. What does PersonX likely **want** to do next **after** this event?

3. Does this event affect people other than PersonX?

(e.g., PersonY, people included but not mentioned in the event)

Yes  No

a). What do they likely **want** to do next **after** this event?

# Quasimodo evaluation data

- Not a KB construction effort!
- Only tiny slice of humans data for evaluation

<https://www.mpi-inf.mpg.de/fileadmin/inf/d5/research/quasimodo/CSK-crowd-for-recall.xlsx>

## 2400 statements

Surgeons are the ones who operate	Surgeons are medical specialists	Surgeons work in the operating rooms	Surgeons have a good salary
Surgeons are precise.	Surgeons have studied.	Surgeons work in hospitals.	Surgeons work under sterile conditions.
surgeons are a physicians who performs	surgeons work in hospitals.	surgeons work also in podiatry.	surgeons work also in dentistry and the
a surgeon is a specialist doctor	a surgeon works with his hands	a surgeon is able to prevent and cure dis	To be a surgeon you need a specializati
Surgeons perform complex operations	Surgeons treat people	Surgeons work in clinics	Brave surgeons
surgeons save people's lives	the surgeons are real heroes	surgeons work many hours a day	surgeons must always be available

[Romero et al., CIKM 2019]

We are collecting common knowledge, where you are given a subject, and you should tell us 4 simple English general sentences about that subject that quickly come to your mind.

## Examples

**Subject: Elephants**

Some possible general sentences:

- *Elephants are grey.*
- *Elephants live in Africa.*
- *Elephants have big ears.*
- ....

**Subject: Nurses**

Some possible general sentences:

- *Nurses work in hospitals .*
- *Nurses are compassionate .*
- *Nurses wear white.*
- ...

## Rules

1. Write full grammatical sentences that start with the subject ("*Elephants are grey.*", not "*They are grey.*").
2. Sentences should contain verbs ("*Elephants have trunks.*"), not just be word clouds (Not "*Trunk*" nor "*Elephant trunk*").
3. Sentences should be general ("*Elephants are big*"), not refer to your personal situation ("*I saw an elephant last week*").
4. **Submissions not adhering to these rules will be rejected!**

Still lots of misunderstandings

“I know a surgeon”

“Un cirujano es un gran hombre”

“The heart surgeon”

→ Second round of peer filtering might help

What sentences comes to your mind when you think of **Surgeons**?

**Sentence 1: (required)**

**Sentence 2: (required)**

**Sentence 3: (required)**

**Sentence 4: (required)**

What sentences comes to your mind when you think of **Moths**?

**Sentence 1: (required)**

**Sentence 2: (required)**

**Sentence 3: (required)**

# Wikidata

- Collaborative knowledge base construction effort
- Under umbrella of Wikimedia foundation
- Best public source on encyclopedic knowledge today
- Commonsense:
  - Comparably lower coverage
  - Roughly comparable to ConceptNet
  - Growing...to be monitored

[Ilievski et al, Arxiv 2020]



# Crowdsourcing - Summary

Project	Focus	#statements	Notes
WordNet	Taxonomical relations	175k synsets	Expert-built
Cyc	General statements and rules	OpenCyc: ~2 M statements	Expert-built, closed source
ConceptNet	Object properties	1.6 M statements	
Atomic	Events	877k statements	
Wikidata	Object properties	100k CSK statements	Editable

- Limited by volunteer effort/money
- Targeted domains in reach for industrial efforts
- Quality assurance important

# Outline

1. Introduction to CSK
2. Crowdsourced CSKB construction
- 3. Text-extraction for ACSKB construction**
  - 1. Overview**
  2. Recipe
  3. Example projects
4. CSKBs: Summary and Outlook

# Overview

- **Earliest projects** on CSKB construction were **manually** authored (Cyc, ConceptNet)
- Challenges in scale
  - Atomic: ~100k\$ annotator expenses
- **Automated information extraction** and KB construction field with **long history**
  - Focus traditionally on crisp ``encyclopedic'' knowledge (cf. DBpedia, YAGO, NELL, DeepDive, ...)
- **Can we use automated IE and KBC for CSK?**



Graham Neubig

@gneubig

Following



One commonly cited argument about the difficulty of learning common-sense reasoning is that "no-one writes down common sense". A counter-argument is "well, the web is big": [instructables.com/id/How-To-Open...](https://www.instructables.com/id/How-To-Open...)

## **How to Open a Door**

**Step 1: Locate Desired Door**

**Step 2: Locate Door Handle or Knob**

**Step 3: Turn Knob or Handle and Pull or Push**

# Challenges of automated CSKB construction

- **Underspecified text semantics**
  - “Lions attack humans” – all/some/all the time/once/..?
- **Reporting bias**
  - “woman kills” vs. “woman breathes” – 1.5M vs. 0.1M web search results
  - “pink elephant” vs. “grey elephant” – 6.9M vs. 1.9M web search results
- **Sparse observations of quadratic+ space of possible statements**
  - Do computer programmers drink water?
- **Noise and polysemy**
  - Pigs can fly - idiom
  - Lynx: Constellation, web browser, animal

# Outline

1. Introduction to CSK
2. Crowdsourced CSKB construction
- 3. Text-extraction for ACSKB construction**
  1. Overview
  - 2. Recipe**
  3. Example projects
4. CSKBs: Summary and Outlook

# Recipe – Generic design points

1. Sources
2. Extraction method
3. Type of contextualization
4. Consolidation method

# Design point 1 – Source choices

- *“Where to extract from?”*

- Wikipedia
- Books and other dedicated sources
  - ARC science corpus
  - Project Gutenberg
- Web search
- Forums
  - Reddit
  - Quora
  - Yahoo Answers
- Search engine query logs
- Web crawls
  - ClueWeb
  - CommonCrawl
- ...



Precision  
Coherence

Recall  
Redundancy



# Extraction source - considerations

- (CS)KB projects stand and fall with source selection
- Precision: *Topic-specific sources >> random web*
  - Event knowledge – *Wikihow* [HowToKB, WWW 2017]
  - Cultural knowledge – *Movie scripts* [Knowlywood, CIKM 2015]
  - Science knowledge – *Science textbooks* [GenericsKB, Arxiv 2020]
- Frequency signals may be stronger from general web dumps, but considerable noise
- Intermediate setting: *Targeted web search* [TupleKB, Ascent]

## Design point 2 – Extraction method options

- *“How to extract”*

1. **Manual patterns** [WebChild, WSDM 2014]
  - Hearst patterns etc.
2. **Co-occurrence** [DoQ, ACL 2019]
  - Window, same sentence, ...
3. **Open information extraction** [TupleKB, Quasimodo, Ascent]
  - Any verb phrase
4. Relation-specific **supervised learning**

# Extraction method - considerations

- Preferred method depends on desired knowledge representation
  - E.g.,
    - Few non-overlapping relation → Co-occurrence
    - Moderate relations → Supervised extractors
    - Many relations → OpenIE
- Has implications downstream
  - Extraction confidences (supervised extractors) for quantitative contextualization
  - Text context for qualitative contextualization
  - OpenIE with many unspecific extractions

# Design point 3 – Contextualization

*“What do we annotate statements with?”*

1. Observation frequency [WebChild 2.0, DoQ]
  - *Elephant, has, tusks, 155*
  - *Elephant, has, tail, 84*
2. Quantitative [0,1] truth labels [TupleKB, Quasimodo]
  - *Elephant, lives in, group, 0.87*
3. Qualitative truth labels [Ascent]
  - *Elephant, lives in, group, **temp**: during wet season*
  - ***Subgroup**: Female elephant, lives in, group*

# Contextualization - considerations

- **Frequencies** trivial to interpret, but do not qualify degree of truth
- **Quantitative truth labels** nontrivial semantics
- **Qualitative labels** easier to interpret, but harder to compare
- **Expressive proposals** from KR exist (e.g., modal logics)
  - Actual implementation not easy
    - Sparse realization in natural language
    - Correct extraction nontrivial

# Design point 4 – Consolidation

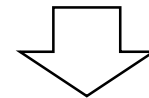
*“What do we do with redundant and competing extractions?”*

- Similar statements may be seen several times
- Redundancy and contradictions may require additional inference
- Common consolidation methods
  1. Keep all [DoQ]
  2. Frequency cutoff [Ascent]
    - E.g., at least seen 5 times
  3. Per-statement consolidation [TupleKB, Quasimodo]
    - Feature-based classification/ranking
  4. Joint consolidation [WebChild, Dice, Ascent]
    - E.g., BERT-based clustering, MaxSAT, ...

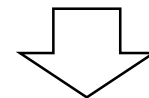
elephant is capable of...

- en carry a trunk →
- en forget to go on the paper →
- en lift logs from the ground →
- en to lift the tree →
- en remember water sources →

*Cats, are, solitary*  
*Lions, live in, groups*



*Lions, are, cats*



# Consolidation - considerations

- Redundancy challenge and blessing
- Exploiting redundancy requires strong text similarity/entailment modules
- Previous projects often stuck to per-statement consolidation due to lack of strong similarity/entailment modules
- Recent advances on pretrained LMs give hope for joint consolidation (see e.g., Dice, Ascent)

# Outline

1. Introduction to CSK
2. Crowdsourced CSKB construction
- 3. Text-extraction for ACSKB construction**
  1. Overview
  2. Recipe
  - 3. Example projects**
4. CSKBs: Summary and Outlook



# Example projects

1. Webchild [Tandon et al., WSDM 2014]
  - Disambiguated noun-adjective pairs
2. **Quasimodo [Romero et al., CIKM 2019]**
  - Salient general triples
3. DoQ [Elazar et al., ACL 2019]
  - Quantitative knowledge
4. Dice [Chalier et al., AKBC 2020]
  - Multifaceted quantitative contextualization and joint consolidation

# Quasimodo

= Query Logs and QA Forums for Salient Commonsense Definitions

- Focus on **salient** knowledge
  - Human associations, curiosity
- **Source:** Query logs and QA forum questions
- **Extraction method:** OpenIE
- **Contextualization:** Supervised precision + IDF
- **Consolidation:** Largely per-statement regression



*(The Hunchback of Notre Dame)*

# Starting point: Humans vs. automated IE

Manual constructions:

- Salient but few

[ConceptNet]

elephant is capable of...

en carry a trunk →

en remember water sources →

(6 more)

Automated construction:

- Many but boring

[TupleKB]

**Elephant:**

- *require, ground*

- *inhabit, region*

- (95 more)

How to reconcile the two?

# Salient knowledge: Utterance context

**Key idea:** Questions convey salient knowledge

- Why do cats purr?
- Why do Americans love guns?
- Why are airplanes white?
  - a) So someone knows these!
  - b) That someone cares enough to ask!

# Salient knowledge: Premier sources

- QA forums:

- Reddit
- Quora
- Yahoo answers
- Ask.com

- Search engine query logs

- Bing
- Google

# Tapping search engine query logs

why do cats

why do cats **purr**

why do cats **like boxes**

why do cats **meow**

why do cats **knead**

why do cats **sleep so much**

why do cats **hate water**

why do cats **like catnip**

why do cats **lick you**

why do cats **have whiskers**

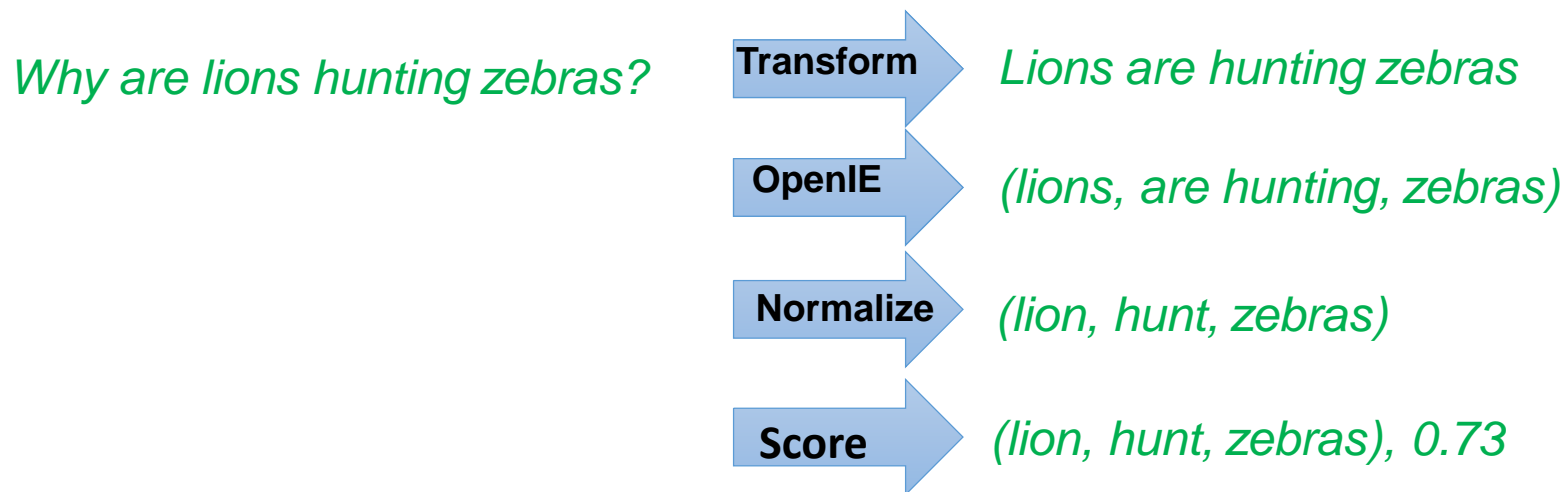
- Autocomplete gives only 10 suggestions/query
  - Exhaustive suffix probing
    - *Why do cats a*
    - *Why do cats b*
    - *Why do cats ...*
    - *Why do cats aa*
    - *Why do cats ab*
    - ...

# Question templates

Pattern	In Query Logs	In QA Forums
how does	19.4%	7.5%
why is	15.8%	10.4%
how do	14.9%	38.07%
why do	10.6%	9.21%
how is	10.1 %	4.31%
why does	8.97%	5.46%
why are	8.68%	5.12%
how are	5.51%	1.8%
how can	3.53%	10.95%
why can't	1.77%	1.40%
why can	0.81%	0.36%

# Statement extraction

- Questions → statements → tuples using OpenIE

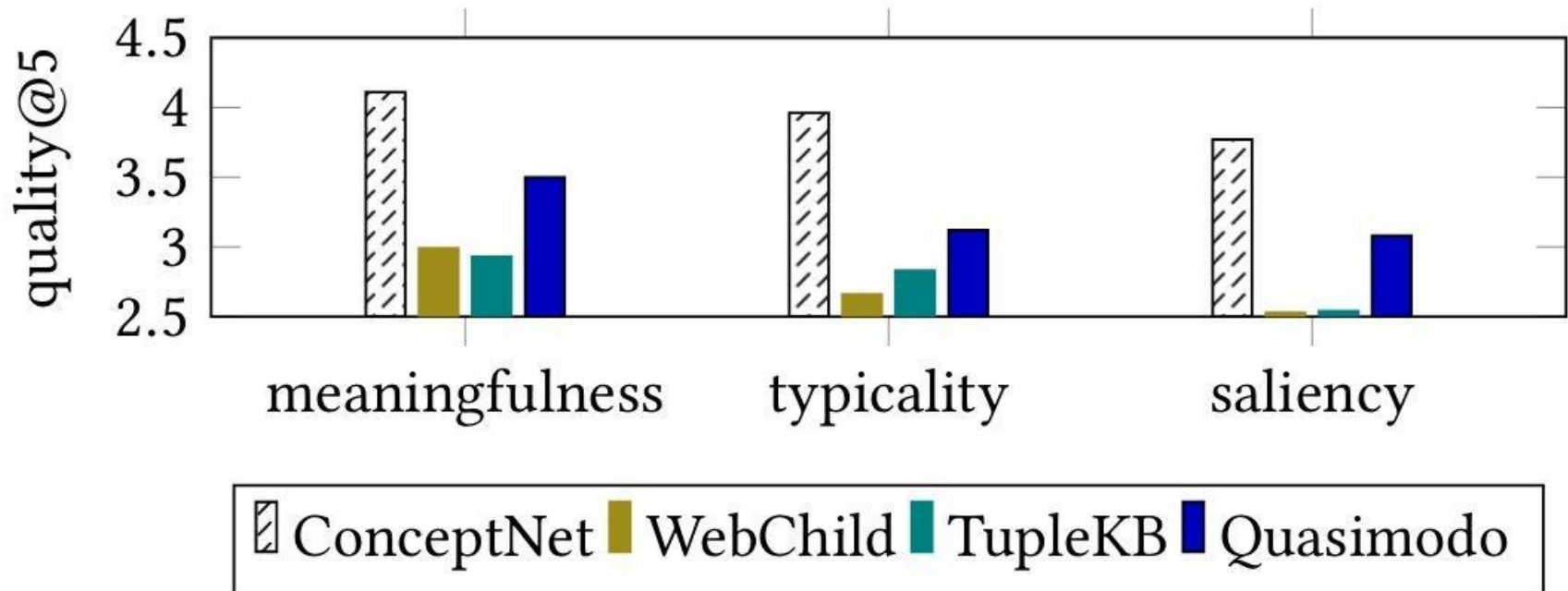




# Anecdotal Examples

Practical human knowledge	<b>(car, slip on, ice)</b>
Problems linked to a subject	<b>(pen, can, leak)</b>
Emotions linked to events	<b>(divorce, can, hurt)</b>
Human behaviors	<b>(ghost, scare, people)</b>
Visual facts	<b>(road, has_color, black)</b>
Cultural knowledge (USA)	<b>(school, have, locker)</b>
Comparative knowledge	<b>(light, faster than, sound)</b>

# Results – Precision



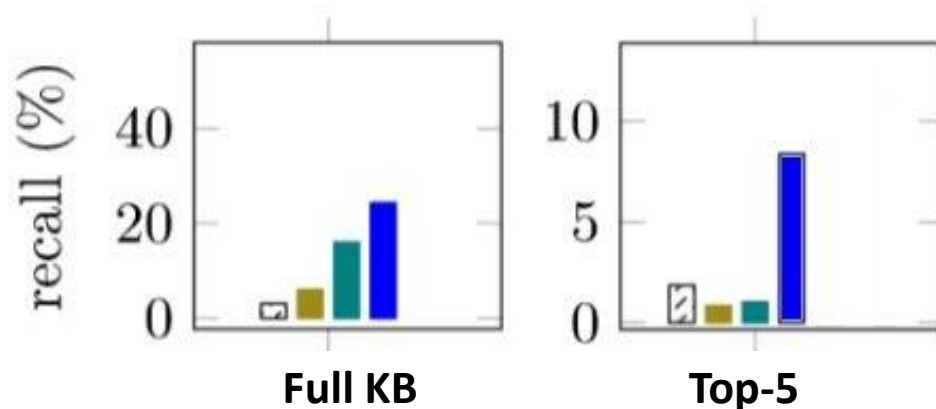
Sample from a list of common subjects (popular animals and occupations)  
5 = best, 1 = worst

# Results – Recall

## Crowd task:

Tell us 3 things that come to your mind when thinking of **lions**.

1. Lions ...
2. Lions ...
3. Lions ...



□ ConceptNet   ■ WebChild   ■ TupleKB   ■ Q'modo

# Extrinsic evaluation

Where would I not want a fox?

👍 hen house, 🗨️ england, 🗨️ mountains,  
🗨️ english hunt, 🗨️ california

KB	Elementary NDMC	Middle NDMC	CommonsenseQA2	Trivia	Examveda	All
#Questions (Train/Test)	623/541	604/679	9741/1221	1228/452	1228/765	10974/3659
Random	25.5	23.7	21.0	25.9	25.4	22.0
word2vec	26.2	28.3	27.8	27.4	25.6	27.2
Quasimodo	<b>38.4</b>	<b>34.8</b>	26.1	<b>28.1</b>	<b>32.6</b>	<b>31.3</b>
ConceptNet	28.5	26.4	<b>29.9 (source)</b>	24.4	27.3	27.5
TupleKB	34.8	25.5	25.3	22.2	27.4	27.5
WebChild	26.2	25.1	25.2	25.9	27.1	24.1

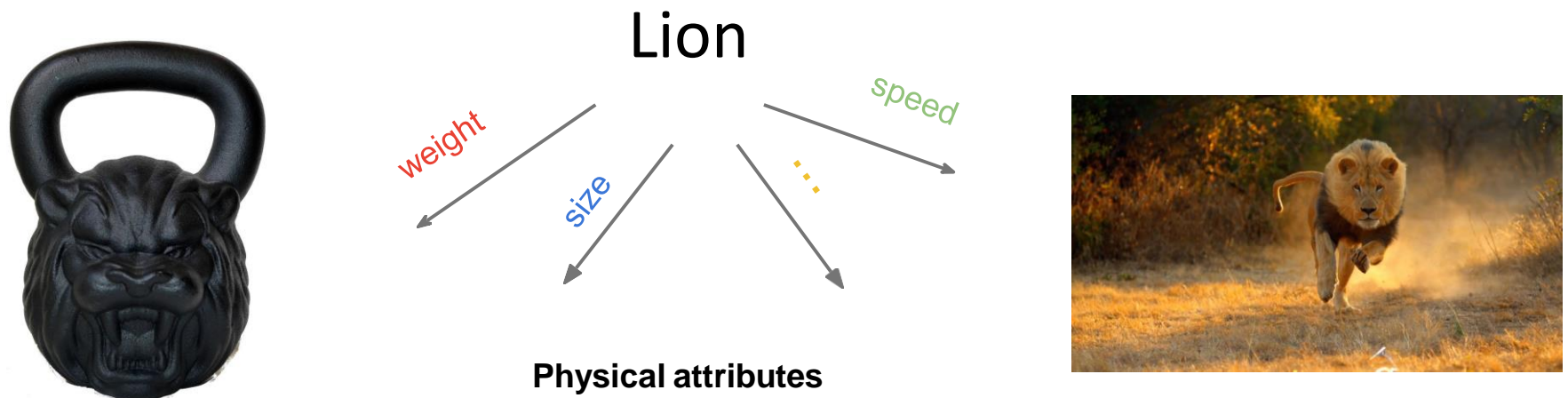
Accuracy in multiple-choice question answering.  
(Simple question-answer connectedness scheme)

# Example projects

1. Webchild 1.0 [Tandon et al., WSDM 2014]
  - Disambiguated noun-adjective pairs
2. Quasimodo [Romero et al., CIKM 2019]
  - Salient general triples
3. **DoQ [Elazar et al., ACL 2019]**
  - Quantitative knowledge
4. Dice [Chalier et al., AKBC 2020]
  - Multifaceted quantitative contextualization and joint consolidation

# Distribution over quantities (DoQ)

- Understanding numerical properties and the way they relate to words.



- Focus on items which can be measured objectively

# Distribution over quantities (DoQ)

- **Source:** Google search engine document index
- **Extraction scheme:** Text window co-occurrence of subject, quantity and dimension keyword
- **Contextualization:** Frequency
- **Consolidation:** none/distribution

# Example - Measurement Detection

*“These breeds can vary in weight from a*

***0.46 kg** teacup poodle ...”*

***Detect numerical measurements using rules:***

***kg/kgs/kilogram -> Mass***

***Normalize (kg -> g)***



# Example - Co-Occurring objects

“These <sup>Noun</sup> breeds can vary in <sup>Noun</sup> weight from a

0.46 kg teacup poodle ...”

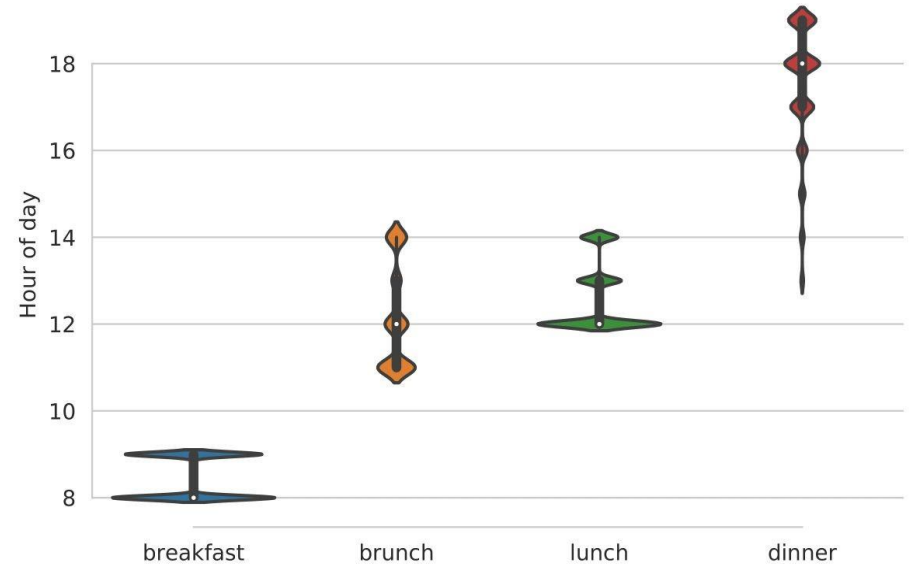
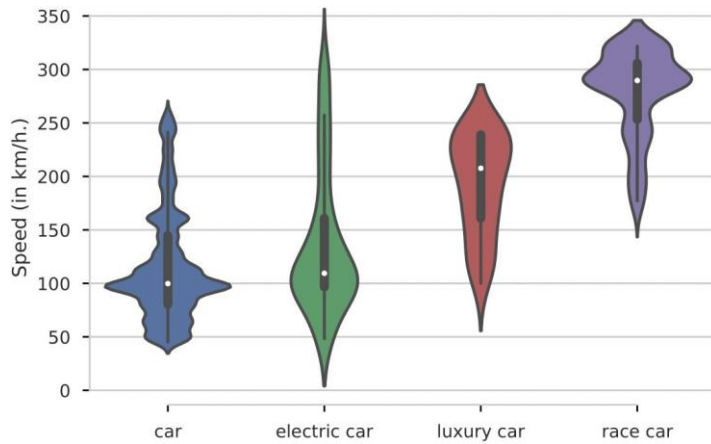
↙

**460 gram**

***Detect objects of interest (Nouns, Adjectives and Verbs) using a POS tagger.***

```
objects_distribution['poodle']['mass'] += [460]
objects_distribution['breeds']['mass'] += [460]
...
```

# Example - Aggregating Measurements



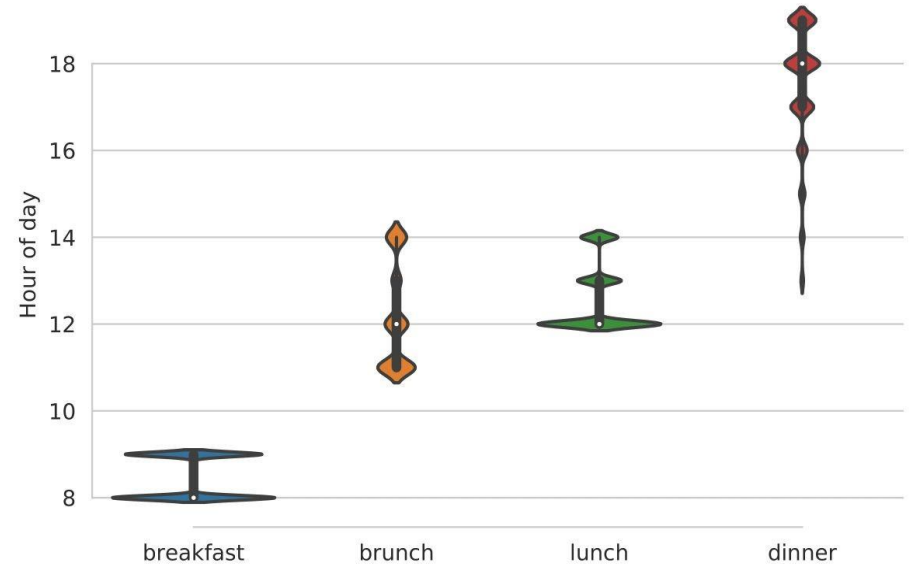
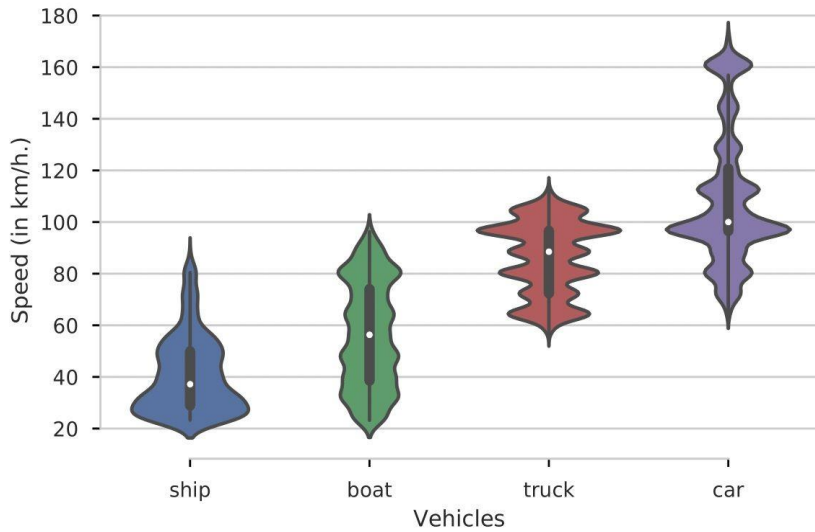
# Resource Statistics - DoQ

- **Distributions over Quantities (DoQ)**
- A very large and diverse resource
- ~120M Unique tuples (object, measurement)
  - ~350K with  $\geq 1000$  occurrences
- Measurement types:
  - Length, mass, currency, temperature, ...
- 27 In total

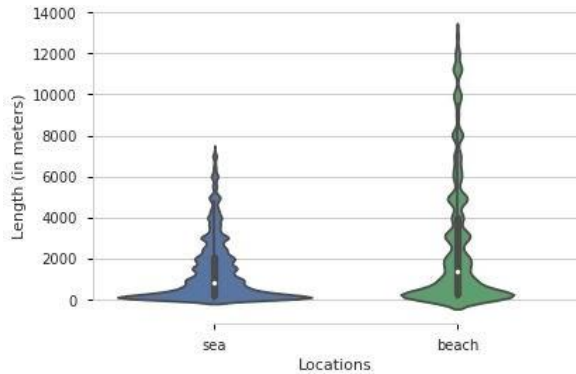
# Intrinsic Evaluation

- Extract the median of “popular” noun distributions
- Expand to a range
  - 20 mm → 10-100 mm
- Ask annotators if the item fits the range
  - “Is the usual length of a screw between 10-100mm?”
- 69% agreement with predictions
- Not perfect, but a reasonable start for acquiring such knowledge

# Comparable Objects - Cool Results



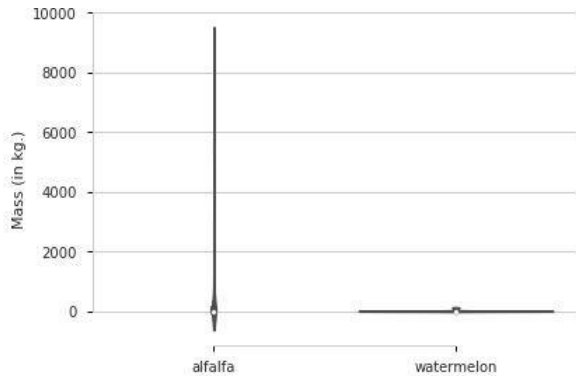
# Comparable Objects - Some Issues



“Elevation ranges from **3,000 feet**  
... above **sea** level.”

That's a small sea!

# Comparable Objects - Some Issues



That's a heavy alfalfa

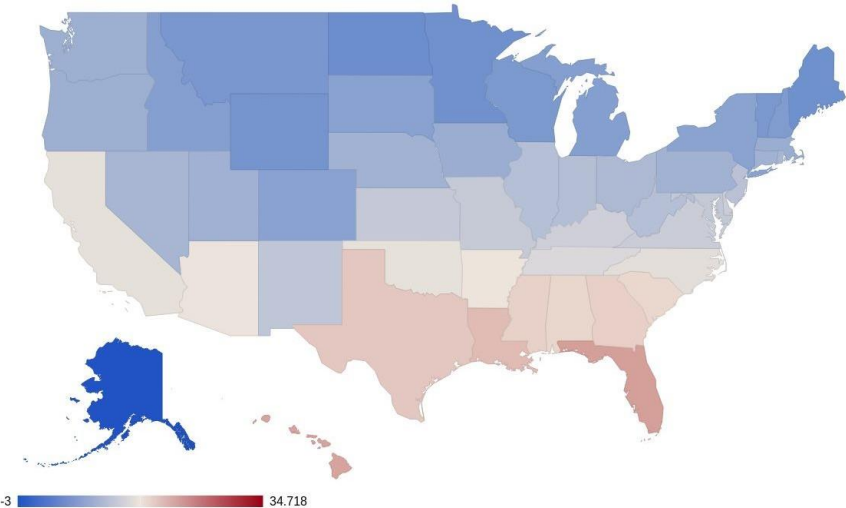
“***Alfalfa*** is the most cultivated legume ... reaching around **454 million tons ...**”



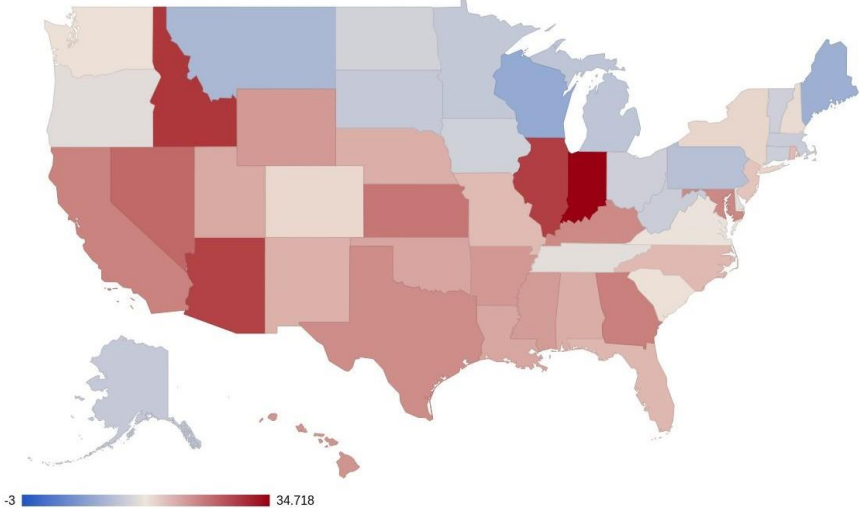
<https://alivebynature.com/the-right-way-to-eat-alfalfa-sprouts/>

# Comparable Objects - Case Study

Collected temperatures of US States



“Real”  
average



Predicted median



# Summary

## 1. Sources

- Domain-specific selection pays off

## 2. Extraction method

- OpenIE vs. trained extractors

## 3. Contextualization

- Expressivity-extractability tradeoff
- Quantitative vs. qualitative

## 4. Consolidation

- Advances in text similarity detection enable joint consolidation

## State of the art

- Automatically extracted CSKBs competitive with manually-built projects
  - Usually huge gains in recall, moderate loss in precision

# Overview – major projects

	Domain	1. Sources	2. Extraction	3. Contextualization	4. Consolidation	Size (#statements)
WebChild	General noun-adjective pairs	Books	Manual patterns	Single precision	Joint ILP	4.6 M
TupleKB	Science triples	Targeted web search	OpenIE	Single precision	Supervised per-statement	0.3 M
Quasimodo	General triples	User questions	OpenIE	Single precision	Supervised per-statement	4 M (v1.3)
DoQ	Quantity triples	Web crawls	Co-occurrence	Frequency	-	(120 M)
Dice	General triples	Existing structured CSKBs	-	Four quantitative facets	Joint MaxSAT	-
Ascent	General triples	Targeted web search	Facet-based OpenIE	Qualitative facets, subject constraints, frequency	Similarity clustering	8.6 M

# Outline

1. Introduction to CSK
2. Crowdsourced CSKB construction
3. Text-extraction for ACSKB construction
  1. Overview
  2. Recipe
  3. Example projects
4. **CSKBs: Summary and Outlook**

# Evaluation

- Intrinsic evaluation

- Size
- Precision
- Saliency
- Recall

→ Based on user judgments/input

- Extrinsic evaluation

- Wide set of academic benchmarks available
  - AllenAI science challenge perhaps most prominent
- Often focus on **reasoning**, not just **knowledge**

# AllenAI challenge

- ~8000 real school questions

*Which property of a mineral can be determined just by looking at it? (A) luster [correct] (B) mass (C) weight (D) hardness*

*A student riding a bicycle observes that it moves faster on a smooth road than on a rough road. This happens because the smooth road has (A) less gravity (B) more gravity (C) less friction [correct] (D) more friction*

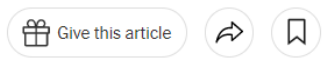
Grade	Challenge		Easy	
	%	(# qns)	%	(# qns)
3	3.6	(94 qns)	3.4	(176 qns)
4	9	(233)	11.4	(591)
5	19.5	(506)	21.2	(1101)
6	3.2	(84)	3.4	(179)
7	14.4	(372)	10.7	(557)
8	41.4	(1072)	41.2	(2139)
9	8.8	(229)	8.7	(454)

Table 2: Grade-level distribution of ARC questions

## Accuracy Over Time



# A Breakthrough for A.I. Technology: Passing an 8th-Grade Science Test



By **Cade Metz**  
Sept. 4, 2019

SAN FRANCISCO — Four years ago, more than 700 computer scientists competed in a contest to build artificial intelligence that could pass an eighth-grade science test. There was \$80,000 in prize money on the line.

They all flunked. Even the most sophisticated system couldn't do better than 60 percent on the test. A.I. couldn't match the language and logic skills that students are expected to have when they enter high school.

But on Wednesday, the Allen Institute for Artificial Intelligence, a prominent lab in Seattle, unveiled a new system that passed the test with room to spare. It correctly answered more than 90 percent of the questions on an eighth-grade science test and more than 80 percent on a 12th-grade exam.

# Benchmarks on commonsense reasoning

- CommonsenseQA
  - <https://www.tau-nlp.org/commonsenseqa>
  - Talmor et al. NAACL, 2019
- CommonGen
  - Lin et al., Arxiv, 2020
  - <https://arxiv.org/abs/1911.03705>
- MC-TACO
  - Zhou et al., EMNLP 2019
  - <https://arxiv.org/pdf/1909.03065.pdf>
- Semeval 2020 Task 4
  - <https://competitions.codalab.org/competitions/21080>
- Further listings:
  - <https://leaderboard.allenai.org/>



# CommonsenseQA

Where would I not want a fox?

👍 hen house, 🗨️ england, 🗨️ mountains,  
🗨️ english hunt, 🗨️ california

Why do people read gossip magazines?

👍 entertained, 🗨️ get information, 🗨️ learn,  
🗨️ improve know how, 🗨️ lawyer told to

What do all humans want to experience in their own home?

👍 feel comfortable, 🗨️ work hard, 🗨️ fall in love,  
🗨️ lay eggs, 🗨️ live forever

# CommonGen


**Concept-Set:** a collection of objects/actions.

dog | frisbee | catch | throw



***Generative Commonsense Reasoning***

**Expected Output:** everyday scenarios covering all given concepts.

- A dog leaps to catch a thrown frisbee. **[Humans]**
- The dog catches the frisbee when the boy throws it.
- A man throws away his dog 's favorite frisbee expecting him to catch it in the air. 

GPT2: A dog throws a frisbee at a football player. **[Machines]**

UniLM: Two dogs are throwing frisbees at each other .

BART: A dog throws a frisbee and a dog catches it.

T5: dog catches a frisbee and throws it to a dog 

# MC-TACO

## Example 1. (event ordering)

*Paragraph:* Growing up on a farm near St. Paul, L. Mark Bailey didn't dream of becoming a judge.

*Question:* What did Mark do right after he found out that he became a judge?

- had a nice dinner
- he buy a 45-acre horse farm
- he celebrated
- mark retired
- he dream of participating himself

## Example 2. (event duration)

*Paragraph:* Growing up on a farm near St. Paul, L. Mark Bailey didn't dream of becoming a judge.

*Question:* How many years did it take for Mark to become a judge?

- 63 years
- 7 weeks
- 7 years
- 7 seconds
- 7 hours

# References

- Based on tutorial @ KI 2020
- Related tutorials
  - Commonsense Reasoning for Natural Language Processing, Sap et al., ACL 2020 (NLP)
  - Common Sense Knowledge Graphs (CSKGs), Ilievski et al, ISWC2020 (Semantic Web)

# References – Major projects

1. Tandon, Niket, et al. "Webchild: Harvesting and organizing commonsense knowledge from the web." *WSDM*. 2014.
2. Mishra, Bhavana Dalvi, Niket Tandon, and Peter Clark. "Domain-targeted, high precision knowledge extraction." *TACL*. 2017
3. Romero, Julien, et al. "Commonsense properties from query logs and question answering forums." *CIKM*. 2019.
4. Elazar, Yanai, et al. "How large are lions? inducing distributions over quantitative attributes." *ACL*. 2019
5. Chalier, Yohan, et al. "Dice: A Joint Reasoning Framework for Multi-Faceted Commonsense Knowledge" *AKBC*. 2020
6. Nguyen, Tuan-Phong, Simon Razniewski, and Gerhard Weikum. "Advanced Semantics for Commonsense Knowledge Extraction." *WWW* 2021.

# Further references

- Omeliyanenko, Janna, et al. "LM4KG: Improving Common Sense Knowledge Graphs with Language Models." *ISWC*, 2020.
- Bhakthavatsalam, Sumithra, Chloe Anastasiades, and Peter Clark. "GenericsKB: A Knowledge Base of Generic Statements." *arXiv preprint arXiv:2005.00660* (2020).
- Bhakthavatsalam, Sumithra, et al. "Do dogs have whiskers? a new knowledge base of haspart relations." *arXiv preprint arXiv:2006.07510* (2020).
- Tandon, Niket, et al. "Knowlywood: Mining activity knowledge from hollywood narratives." *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. 2015.
- Chu, Cuong Xuan, Niket Tandon, and Gerhard Weikum. "Distilling task knowledge from how-to communities." *Proceedings of the 26th International Conference on World Wide Web*. 2017.
- Schubert, Lenhart. "Can we derive general world knowledge from texts." *Proc. HLT*. 2002
- Schubert, Lenhart, and Matthew Tong. "Extracting and evaluating general world knowledge from the Brown corpus." *Proceedings of the HLT-NAACL 2003 workshop on Text meaning*. 2003.

# Take home

1. Structured CSK important **interpretable and scrutable building block** for trustworthy AI
2. **Coherence and density require consolidation**
  1. Multi-source validation
  2. Constraint-based reasoning
3. **Semantics of CSK** still with gaps
  1. Opportunity for deliberate KR
4. Advance of **neural models** suggest **hybrid architectures**
  1. Neural model for bridging language gaps (see also next lecture)