max planck institut
informatik

Beyond Knowledge Graphs: New Frontiers of Machine Knowledge

Gerhard Weikum

Max Planck Institute for Informatics

http://mpi-inf.mpg.de/~weikum

Outline

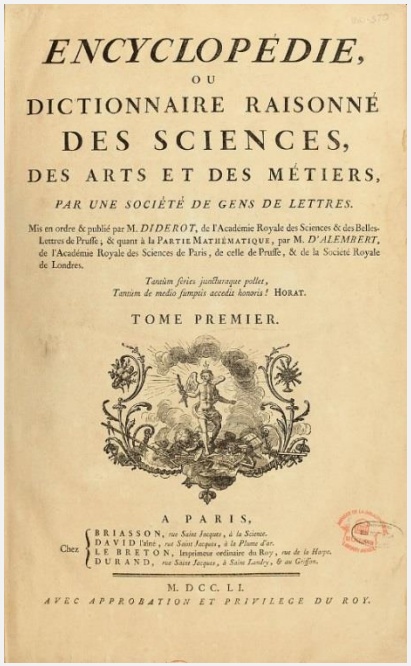
- ★ 1 Knowledge Graphs …
- ★ 2 … And Beyond
- ★ Conclusion

Brief History of Knowledge Bases

Didier Diderot

Jean Le Rond d‘Alembert

1751-1780



ENCYCLOPÉDIE,
OU
DICTIONNAIRE RAISONNÉ
DES SCIENCES,
DES ARTS ET DES MÉTIERS,
PAR UNE SOCIÉTÉ DE GENS DE LETTRES.

Mis en ordre & publié par M. DIDEROT, de l'Académie Royale des Sciences & des Belles-Lettres de Prusse, & quant à la PARTIE MATHÉMATIQUE, par M. D'ALEMBERT, de l'Académie Royale des Sciences de Paris, de celle de Prusse, & de la Société Royale de Londres.

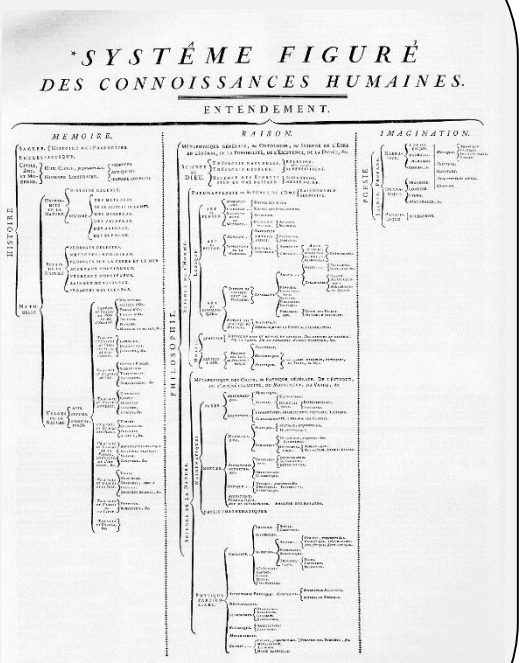
Tantùm series juncturaque pollet,
Tantùm de medio sumptis accedit honoris! HORAT.

TOME PREMIER.

A PARIS,

Chez BRIASSON, rue Saint Jacques, à la Science.
DAVID l'aîné, rue Saint Jacques, à la Plume d'or.
LE BRETON, Imprimeur ordinaire du Roy, rue de la Harpe.
DURAND, rue Saint Jacques, à Saint Landry, & au Griffon.

M. DCC. LI.
AVEC APPROBATION ET PRIVILEGE DU ROY.

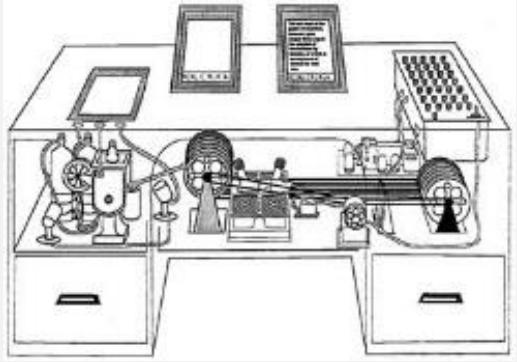


72000 articles, 2250 contributors

1700 1800 1900 1950 1990 1995 2005 2010 2015

Brief History of Knowledge Bases

Vannevar Bush
1945



… a growing mountain of research …
a memex is a … supplement to memory

Denis Diderot

Jean d'Alembert

1700 1800 1900 1950 1990 1995 2005 2010 2015

Brief History of Knowledge Bases

Cyc

Douglas Lenat 1984

$\forall$ x: human(x) $\Rightarrow$
($\exists$ y: mother(x,y) $\wedge$
$\exists$ z: father(x,z))

$\forall$ x,u,w: (mother(x,u) $\wedge$
mother(x,w)
$\Rightarrow$ u=w)

Vannevar Bush

Denis Diderot

Jean d'Alembert

1700 1800 1900 1950 1990 1995 2005 2010 2015

Brief History of Knowledge Bases

WordNet

George Miller

Christiane Fellbaum

1985

guitarist ⊂ {player,musician} ⊂ artist

{player,footballer} ⊂ athlete

Vannevar Bush

Denis Diderot

Jean d'Alembert

1700 1800 1900 1950 1990 1995 2005 2010 2015

Brief History of Knowledge Bases

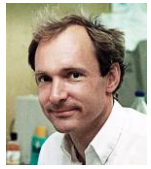
Denis Diderot

Jean d'Alembert

Vannevar Bush

Cyc

WordNet



WWW

Wikipedia

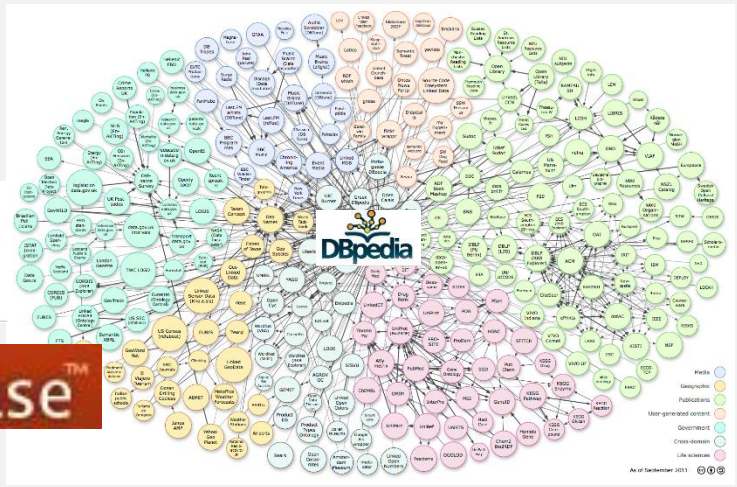
1700 1800 1900 1950 1990 1995 2005 2010 2015

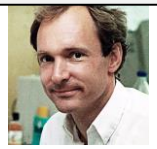
Brief History of Knowledge Bases

yago
select knowledge

DBpedia

freebase™





WWW

Vannevar Bush

Cyc

Denis Diderot

Jean d'Alembert

1700 1800 1900 1950 1990 1995 2005 2010 2015

Brief History of Knowledge Bases

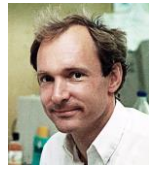
Denis Diderot

Jean d'Alembert

Vannevar Bush

Cyc

WordNet



WWW

Wikipedia

freebase

DBpedia

yago
select knowledge

WIKIDATA

BabelNet

IBM Watson

WolframAlpha

bing

amazon

Alibaba.

SIEMENS
Ingenuity for

Bloomberg

1700 1800 1900 1950 1990 1995 2005 2010 2015

Encyclopedic Machine Knowledge

subject-predicate-object triples about entities, attributes of and relations between entities

plus reification

predicate (subject, object)

---

type (SteveJobs, entrepreneur)
subtypeOf (entrepreneur, businessperson)

taxonomic knowledge

hasFounded (SteveJobs, Apple)
hasDaughter (SteveJobs, LisaBrennan)
diedOf (SteveJobs, pancreatic cancer)
diedOn (SteveJobs, 5-Oct-2011)

factual knowledge

hasSymptom (pancreatic cancer, jaundice)
treats (ErlotinibHydrochloride, pancreatic cancer)

expert knowledge

foundingEvent (123, blank)
founder (123, SteveJobs)
foundedOrg (123, Apple)
foundedOn (123, 1-April-1976)
foundedAt (123, LosAltos-CA-USA)

spatial & temporal knowledge

Mio‘s of entities & Bio‘s of facts
→ semantic search & QA
→ language understanding
→ text analytics, data cleaning …

Machine Knowledge for Answer Engines

Precise and concise answers for advanced information needs:

properties of entity

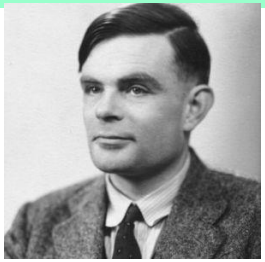
★ **Nobel laureate whose daughter also won a Nobel prize?**

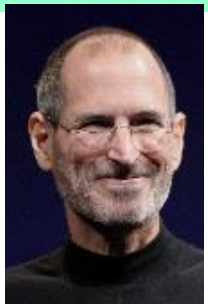
★ **Pop singers who are also poets?**

sets of entities

relationships between entities

★ **Commonalities & relationships among:**
Alan Turing, Paul Bocuse, Steve Jobs, Katherine Goble?





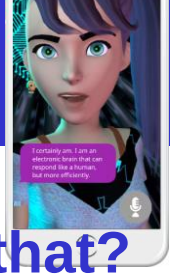
Machine Knowledge for Answer Engines

Precise and concise answers
for advanced information needs:

real applications

- **Proteins that bind to the Zika virus?**
  Polymer materials for super-capacitators?
  European politicians mentioned in Panama Leaks?

Conversational Chatbots



My name is Jerry, cool to meet you Mitsuku.

Cool is a nice middle name. Why did your parents call you that?

What is the highest mountain on earth?

Mt. Everest. It‘s 8848 metres high.

Isn‘t the K2 higher?

A large mountain in the Himalayas.

How high is the Eiffel tower?

It is about 324 metres.

And in feet?

Oh I don‘t know. We use the metric system now.

Which is higher the Eiffel tower or the Everest?

The one closest to the sky.

What is taller a giraffe or a cat?

Are you serious, a giraffe is a lot bigger than a cat.

What is longer, a king cobra or a worm?

The cobra as it has 8 letters.

+ Language Skills

+ Encyclopedic Knowledge

? Quantities

? Commonsense

https://pandorabots.com/mitsuku/

Visual AI: from Perception to Cognition

Computer Vision → Image Description → Question Answering

man in blue wetsuit
is surfing on a wave

woman is holding
a bunch of bananas

Source: https://cs.stanford.edu/people/karpathy/deepimagesent/

Who is wearing glasses?

man

woman

What is the
polar bear eating?

carrot

Source: http://www.visualqa.org/

? Quantities

? Commonsense

What is behind the giraffe?
→ mountain

What is on the mountain peak?
→ snow

What is the largest object
in this image?
→ giraffe

How many legs does
the giraffe have?
→ 3

Outline

- ★ 1 Knowledge Graphs …
- ★ 2 … And Beyond
  - ★ Quantities
  - ★ Commonsense
- ★ Conclusion

Quantities: More than just Numbers

Counts	**1200 submissions, 200 accepted**
Fractions	**5% of the world population, 30% of the guns**
Rates	**FB increased revenue by 30%**
Dates	**24-April-2018, today, last weekend, Easter**
Money	**acquired for 80 Mio USD**
Time	**for 3 years, new record 9.58 s**
Scores	**Brazil vs. Germany 1:7**
Ranks	**4th in the Olympics final**
Physical Measures	**407 MW renewable energy**
	3400 kWh annual energy consumption
Medical Measures	**15mg Xarelto daily**
	Potassium low: 2 mmol/L

Quantities: Measure, Value, Unit

Quantity = (Measure, Value, Unit)

normalize: scale & precision

(Distance, 200, km)	2E5: scale 10^5 prec 1
(Height, 1.80, m)	1.80E0: scale 10^0 prec 3
(Power, 407, MW)	4.07E8: scale 10^8 prec 3
(Energy, 3400, kWh)	3.4E6: scale 10^6 prec 2
(Price, 80,000,000, USD)	8.0E7: scale 10^7 prec 2

Quantity meta-properties:

- typical / normal range
- value distribution
- unit conversions

Reference systems:

SI: International System of Units

UNECE: Codes for Units of Measurement used in the International Trade

Kilowatt hour

Residential electricity meter located in Canada

Unit information	
Unit system	Non-SI metric
Unit of	Energy
Symbol	kW·h

Unit conversions	
1 kW·h in ...	*... is equal to ...*
SI units	3.6 MJ
CGS units	3.6×10^{13} erg
English Engineering units	≈ 2,655,224 ft·lbf
British Gravitational units	≈ 85,429,300 ft·pdl

Quantities in Knowledge Bases

WIKIDATA

Usain Bolt (Q1189)

height	195 centimetre 1 reference
mass	94 kilogram 1 reference

Tesla Model X (Q1634161)

height	1,684 millimetre 1 reference
width	1,999 millimetre 1 reference

SR/OLYMPIC SPORTS

Men's 100 metres Event History · Glossary · SHARE · Embed · CSV · Export · PRE

Games	Age	City	Sport	Country	Phase	Unit	Rank		T(A)	RT
2008 Summer	21	Beijing	Athletics	Jamaica	Final		1	WR	9.69	0.165
2008 Summer	21	Beijing	Athletics	Jamaica	Semi-Finals	Heat One	1	Q	9.85	0.161
2008 Summer	21	Beijing	Athletics	Jamaica	Quarter-Finals	Heat Four	1	Q	9.92	0.165
2008 Summer	21	Beijing	Athletics	Jamaica	Round One	Heat One	1	Q	10.20	0.186
2012 Summer	25	London	Athletics	Jamaica	Final		1		9.63	0.165
2012 Summer	25	London	Athletics	Jamaica	Semi-Finals	Heat Two	1	Q	9.87	0.180
2012 Summer	25	London	Athletics	Jamaica	Round One	Heat Four	1	Q	10.09	0.178
2016 Summer	29	Rio de Janeiro	Athletics	Jamaica	Final		1		9.81	0.155
2016 Summer	29	Rio de Janeiro	Athletics	Jamaica	Semi-Finals	Heat Two	1	Q	9.86	0.143
2016 Summer	29	Rio de Janeiro	Athletics	Jamaica	Round One	Heat Seven	1	Q	10.07	0.156

Men's 200 metres Event History · Glossary · SHARE · Embed · CSV · Export · PRE

Games	Age	City	Sport	Country	Phase	Unit	Rank		T(A)	RT	L
2004 Summer	17	Athina	Athletics	Jamaica	Round One	Heat Four	5		21.05	0.254	5
2008 Summer	21	Beijing	Athletics	Jamaica	Final		1	WR	19.30	0.182	
2008 Summer	21	Beijing	Athletics	Jamaica	Semi-Finals	Heat Two	1	Q	20.09	0.175	
2008 Summer	21	Beijing	Athletics	Jamaica	Quarter-Finals	Heat One	1	Q	20.29	0.186	
2008 Summer	21	Beijing	Athletics	Jamaica	Round One	Heat Five	2	Q	20.64	0.177	
2012 Summer	25	London	Athletics	Jamaica	Final		1		19.32	0.180	
2012 Summer	25	London	Athletics	Jamaica	Semi-Finals	Heat Two	1	Q	20.18	0.192	
2012 Summer	25	London	Athletics	Jamaica	Round One	Heat One	1	Q	20.39	0.191	
2016 Summer	29	Rio de Janeiro	Athletics	Jamaica	Final		1		19.78	0.156	
2016 Summer	29	Rio de Janeiro	Athletics	Jamaica	Semi-Finals	Heat Two	1	Q	19.78	0.156	
2016 Summer	29	Rio de Janeiro	Athletics	Jamaica	Round One	Heat Nine	1	Q	20.28	0.177	

2016 Tesla Model X fuel economy and operating costs

Model	Model year	Fuel efficiency (MPGe)		
		Combined	City	Highway
AWD 90D (90 kWh)[12]	2016	92; 34 kWh/100 mi or 21 kWh/100 km	90; 37 kWh/100 mi or 23 kWh/100 km	94; 32 kWh/100 mi or 20 kWh/100 km
AWD P90D (90 kWh)[47][48]	2016	89; 38 kWh/100 mi or 24 kWh/100 km	89; 38 kWh/100 mi or 24 kWh/100 km	90; 38 kWh/100 mi or 24 kWh/100 km
AWD P100D (100 kWh)	2016	86; 38 kWh/100 mi or 24 kWh/100 km	81; 38 kWh/100 mi or 24 kWh/100 km	92; 38 kWh/100 mi or 24 kWh/100 km

Quantities in Relations

[Sarawagi et al.: AAAI'16,
Mausam et al.: ACL'17,
Alonso et al.: ICWE'15,
Polleres et al.: ISWC'16
Ibrahim et al.: CIKM'16]

Extracting quantities from text & tables

Bolt won the 200m Olympics in 2008 in 19.30s
Bolt won 100 meters Gold in Rio in 9.81 seconds
Xarelto is usually applied after thromboses at 20mg twice per day
Xarelto for children should be limited to a daily dosage of 5-15 mg

→ **hasWon (UsainBolt, OlympicGold, 2008, 200m, 19.30s)**
hasWon (UsainBolt, OlympicGold, 2016, 100m, 9.81s)
treatment (Xarelto, thrombosis, adults, 40mg/day)
treatment (Xarelto, thrombosis, children, 15mg/day)

Problems:

- **n-ary relations expressed in more complex language**
- **partial observations of facts from multiple sentences**

Quantities in Relations

Extracting tuples with quantities from text & tables

[P. Ernst et al.: WWW'18]

Problems:

- **n-ary relations expressed in more complex language**
- **partial observations of facts from multiple sentences**

Bolt won the 200m Olympics in 19.30 seconds
Bolt won second Gold in 2008 with 19.30s record
Bolt ran the 100m final in Rio in 9.81
Bolt won 100m Gold in Rio

→ **hasWon (UsainBolt, OlympicGold, ?X, 200m, 19.30s)**
hasWon (UsainBolt, OlympicGold, 2008, ?Y, 19.30s)
hasWon (UsainBolt, ?W, 2016, 100m, 9.81s)
hasWon (UsainBolt, OlympicGold, 2016, 100m, ?Z)

Solution: use Weighted MaxSat for constraint reasoning over partially grounded fact candidates

Quantity Knowledge: Challenges

Representation:

- **more triples, n-ary relations, nested tuples?**
  emission (TeslaS, California, 2016, measure(0.25, kg/km, CO2))
- **meta-properties?**
  max(height(giraffe),5m)
  unit (distance(city,city),km), unit (distance(star,star),lightyears)

Acquisition:

- **human experts, crowdsourcing, learning from tables/stats?**
- **integrate existing reference systems (SI, medicine, finance, …)**

Usage:

- **search with quantities**
  car models with emission below 0.2 kg/km CO2
- **aggregation & comparison for analytics**
  world-wide avg. dosage of beta blockers per age group

Outline

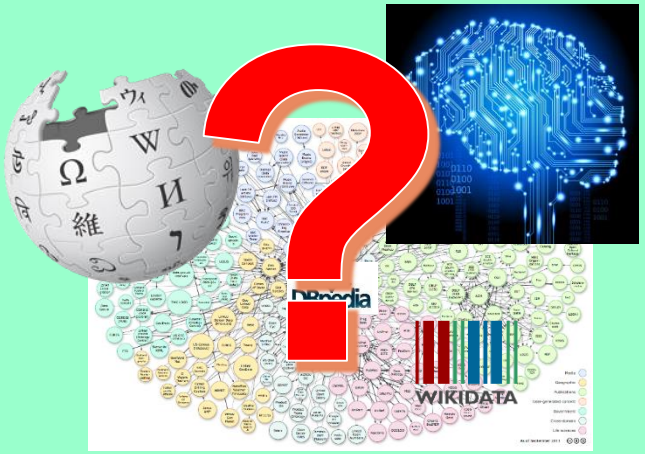
- ★ 1 Knowledge Graphs …
- ★ 2 … And Beyond
  - ★ Quantities
  - ★ Commonsense
- ★ Conclusion

Commonsense Knowledge:
Not So Common

Every child knows that

apples are green, red, round, juicy, …
but not fast, smart, funny, verbose, …

children live with their parents
programmers need computers
… … …



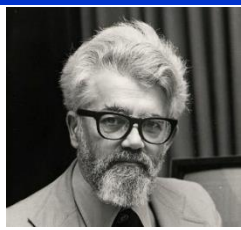
But: commonsense is rarely stated explicitly
reporting in web and social media is sparse and biased

spider 3 legs: 7 Mio
spider 8 legs: 1 Mio

singer: 236 Mio
programmer: 40 Mio
manager: 570 Mio

pink elephant: 16 Mio
grey elephant: 9 Mio

Commonsense Knowledge Bases



John McCarthy
(Turing Award
1971)

"Commonsense knowledge is key to generality in AI"

- **Properties of Everyday Objects**
- **Human Activities / Abilities / Emotions**
- **Plausibility Invariants (incl. Causality …)**
- **Class/Concept Taxonomy**

Cyc

Douglas Lenat

$\forall$ x: human(x) $\Rightarrow$
($\exists$ y: mother(x,y) $\wedge$
$\exists$ z: father(x,z))

$\forall$ x,u,w: (mother(x,u) $\wedge$
mother(x,w)
$\Rightarrow$ u=w)

WordNet

George Miller

Christiane Fellbaum

guitarist $\subset$ {player,musician} $\subset$ artist

{player,footballer} $\subset$ athlete

Acquiring Commonsense Knowledge

Approach 1: Knowledge engineers

→ **Cyc, WordNet**

Problem: scope and scale

Approach 2: Crowdsourcing

→ **ConceptNet [Speer&Havasi],**
VisualGenome [Krishna et al.], …

Problem: scale and quality

Approach 3: Information Extraction

→ **WebChild [Tandon et al.],**
AI2 Alexandria [Etzioni et al.]

Problems: noise and bias

many specific approaches for subclassOf (hypernymy)

WebChild


[N. Tandon & G. de Melo et al.: WSDM'14, AAAI'16, ACL'17]

Use seeds

hasProperty (apple,round), hasAbility (fish,swim), hasLoc (fish,water)

to learn patterns

X is very Y, X can Y, X is in Y, …

for refined properties hasLoc, hasColor, hasShape, hasSize, hasTaste …

Overcome sparseness and bias by
tapping into books (n-grams) and image tags

hasColor (elephant, grey), hasShape (circle, round)
hasAbility (fish, bite), hasAbility (human, talk) …
usedFor (book, learn), usedFor (computer, learn) …
partOf (wheel, bike), partOf (wheel, car) …
hasTemp (weather, crisp), hasTaste (cookie, crisp) …
… … …

2 Mio concepts
18 Mio. assertions
for 30 relations
- semantically typed
- sense disambiguated

https://gate.d5.mpi-inf.mpg.de/webchild/

WebChild: Limitations

Highly varying confidence:

hasColor (apple, …): red, green, yellow, pink, purple, black, white …

hasTaste (apple, …): sweet, sour, juicy, chewy, healthy, cheap …

hasProperty (mountain, …): high, steep, cold, lethal, far, holy, true …

Consider epistemic logic with modalities:
always, usually, possibly, never

Subjective properties:

hasAppearance (kangaroo, cute)

hasProperty (spider, dangerous)

hasProperty (grasshopper, edible), hasProperty (FoieGras, edible)

Consider belief logics:

fromFrance(x) ⇒ belief (x, hasProperty (FoieGras, edible))

¬ fromAustralia(x) ⇒ belief (x, hasAppearance (kangaroo, cute))

Human Activities & Behaviors

[N. Tandon et al.: WWW'15, CIKM'15, WWW'17]

climb mountain	
Agent:	…
Participant:	…
Object:	…
Location:	…
Time:	…

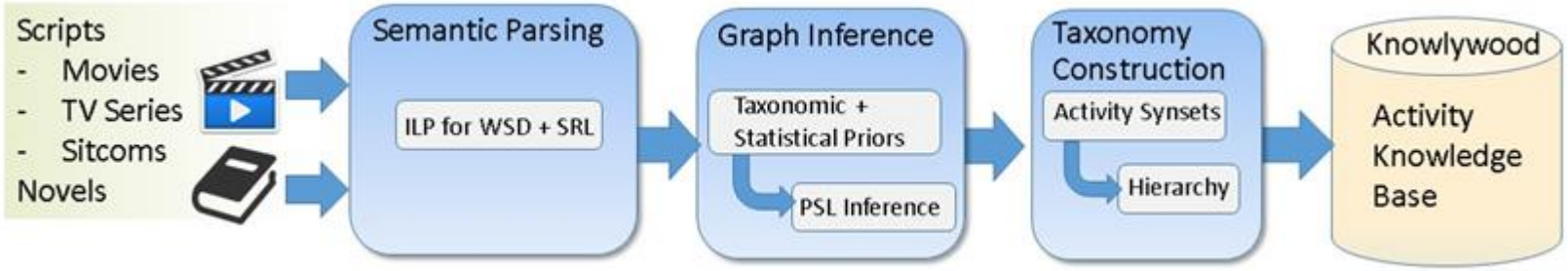
Type

reach summit	
Agent:	…
Participant:	…
Object:	…
Location:	…
Time:	…

Next

climb on ice	
Agent:	mountaineer, man, woman
Participant:	guide, sherpa
Object:	rope, crampons, oxygen mask
Location:	outdoor, mountain peak
Time:	daytime, sunshine

Tap narratives & movie scripts → Knowlywood project



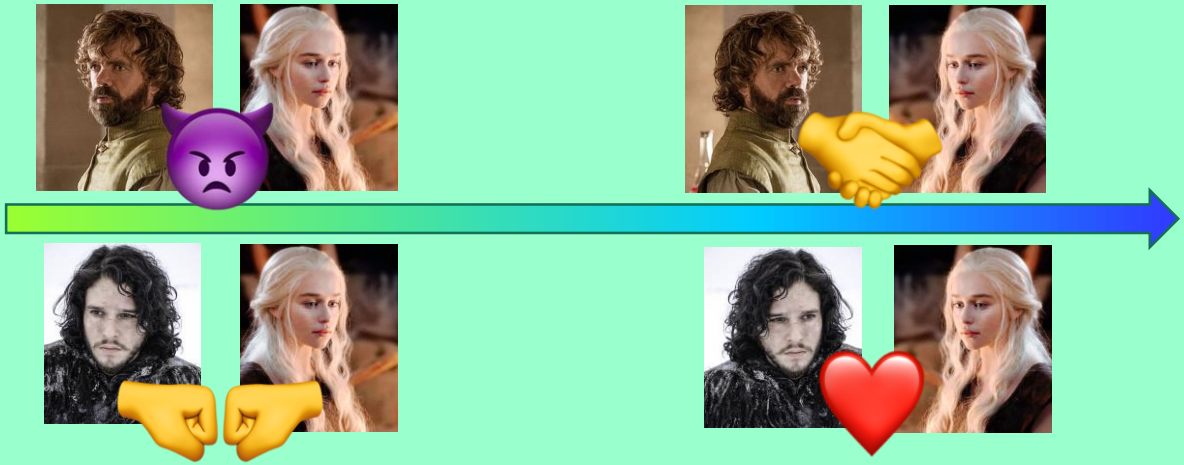
https://gate.d5.mpi-inf.mpg.de/webchild/

Human Emotions


[H. Jhavar & Paramita: WWW'18]

I got invited to give a talk at the Wiki workshop

I got invited to this stupid dinner with my boss



Plutchik's wheel of emotions

Source:
https://www.cheatography.com/davidpol/cheat-sheets/plutchik-s-wheel-of-emotions/

HappyDB:
100,000 happy moments [Asai et al.: LREC'18] :

I got invited to a friend's wedding.
I went to the park with the kids.
Morning started with chirping of birds.

Socio-Cultural Knowledge

Problem:
Commonsense often seems subjective
Varies with region, age, social group, cultural background

dangerous (snake), dangerous (biking)
dangerous (kangaroo), cute (kangaroo)
edible (grasshopper), edible (FoieGras)
greeting (handshake), greeting (kiss), greeting (wai)

You see a dead cat on the street – what do you do?

A: call the police
B: walk on
C: take a selfie
D: think about Schroedinger‘s cat

Use Case: Visual Understanding

man (o1)

woman (o3)

e-guitar (o2)

drum (o4)

? banana (o5)

+ commonsense knowledge:

hasColor (saxophone, yellow)
hasShape (saxophone, curved)
occursWith (saxophone, guitar)
usedFor (saxophone, music concert)

⇒ saxophone(o5)

Use Case: Visual Understanding

man, guitar, woman, saxophone, drums, …

→ *man playing guitar and woman holding sax*

+ commonsense knowledge: **→ groovy rock concert**

Commonsense Knowledge: Challenges

Representation:

- **triples, n-ary relations, modalities, fuzzy, higher-order?**
  hasTaste(chocolate,sweet,mostly), hasTaste(chocolate, bitter, possibly)
  belief (x,edible(grasshoppers)), belief (y,edible(FoieGras))
- **plausibility invariants (rules)**
  $\forall$ x: human(x) $\Rightarrow$ (($\exists$y: mother(x,y)) $\Rightarrow$ wasPregnant(x))
  $\forall$ x: (animal(x) $\wedge$ hasLegs(x)) $\Rightarrow$ isEven(numLegs(x))

Acquisition:

- **crowdsourcing, learning from online contents?**
- **tap books, movies, videos, social media, games?**

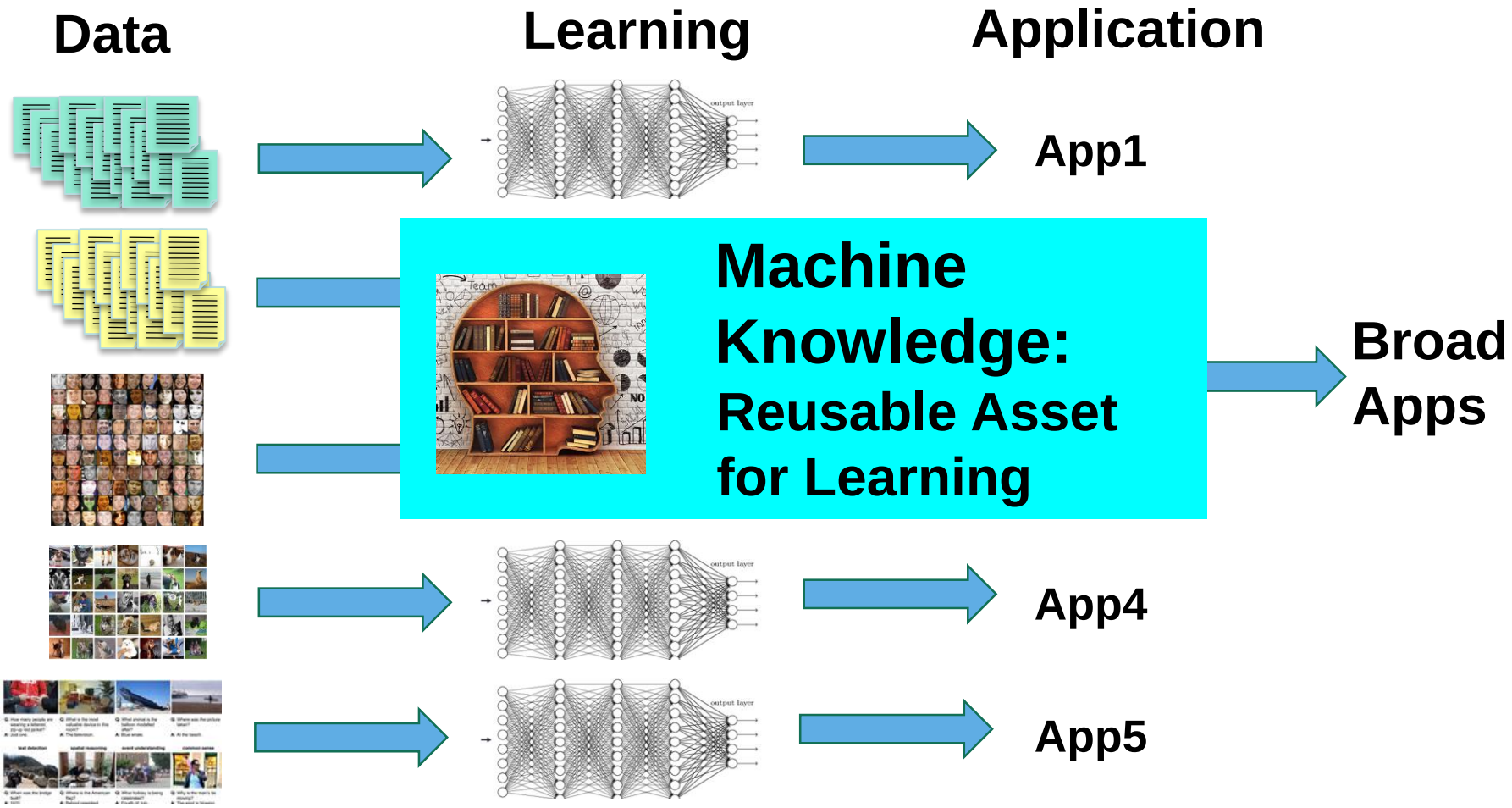
Usage:

- **from narrow chatbots to versatile digital companions**
  (telepresence, games, smart homes …)
- **from visual perception to visual cognition**
  (virtual humans, drones/robots …)

Outline

- ★ 1 Knowledge Graphs …
- ★ 2 … And Beyond
  - ★ Quantities
  - ★ Commonsense
- ★ Conclusion

Deep Learning End-to-End: No Need for Machine Knowledge ?



Knowledge Informs Learning

Take-Home Message

Thank You ! Questions ?

Encyclopedic Knowledge

Entities …

Expert Knowledge

Quantities …

Commonsense Knowledge

Properties …

Socio-Cultural Knowledge

Activities …

Advanced Machine Knowledge: Key Asset for Robust, Interpretable, Versatile AI Applications
(cognitive assistants, visual understanding, …)

Challenges: Representation, Acquisition, Usage