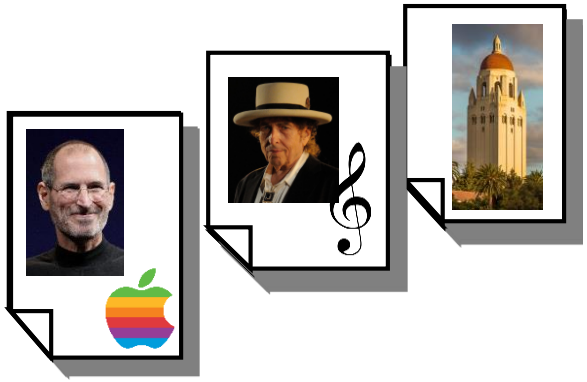
max planck institut
informatik

What Computers Should Know

Gerhard Weikum

Max Planck Institute for Informatics
Saarland Informatics Campus
http://mpi-inf.mpg.de/~weikum

Turn Text & Data into Knowledge



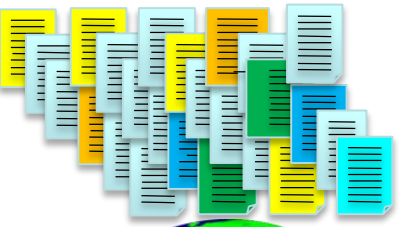
more, deeper, better knowledge

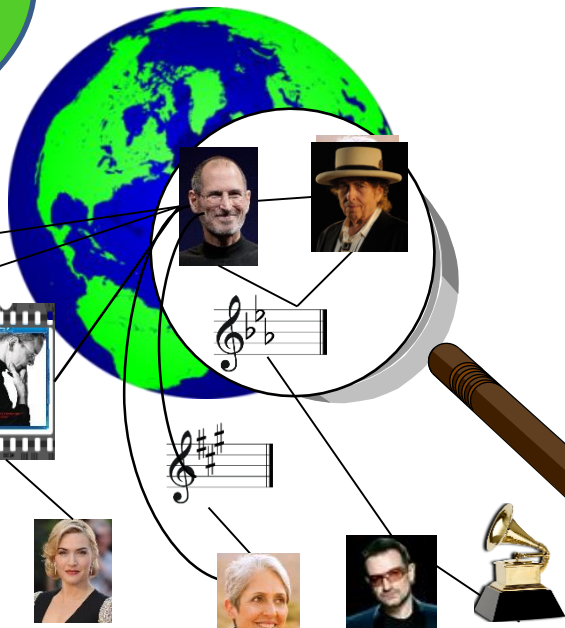
knowledge acquisition

Internet Contents

Knowledge

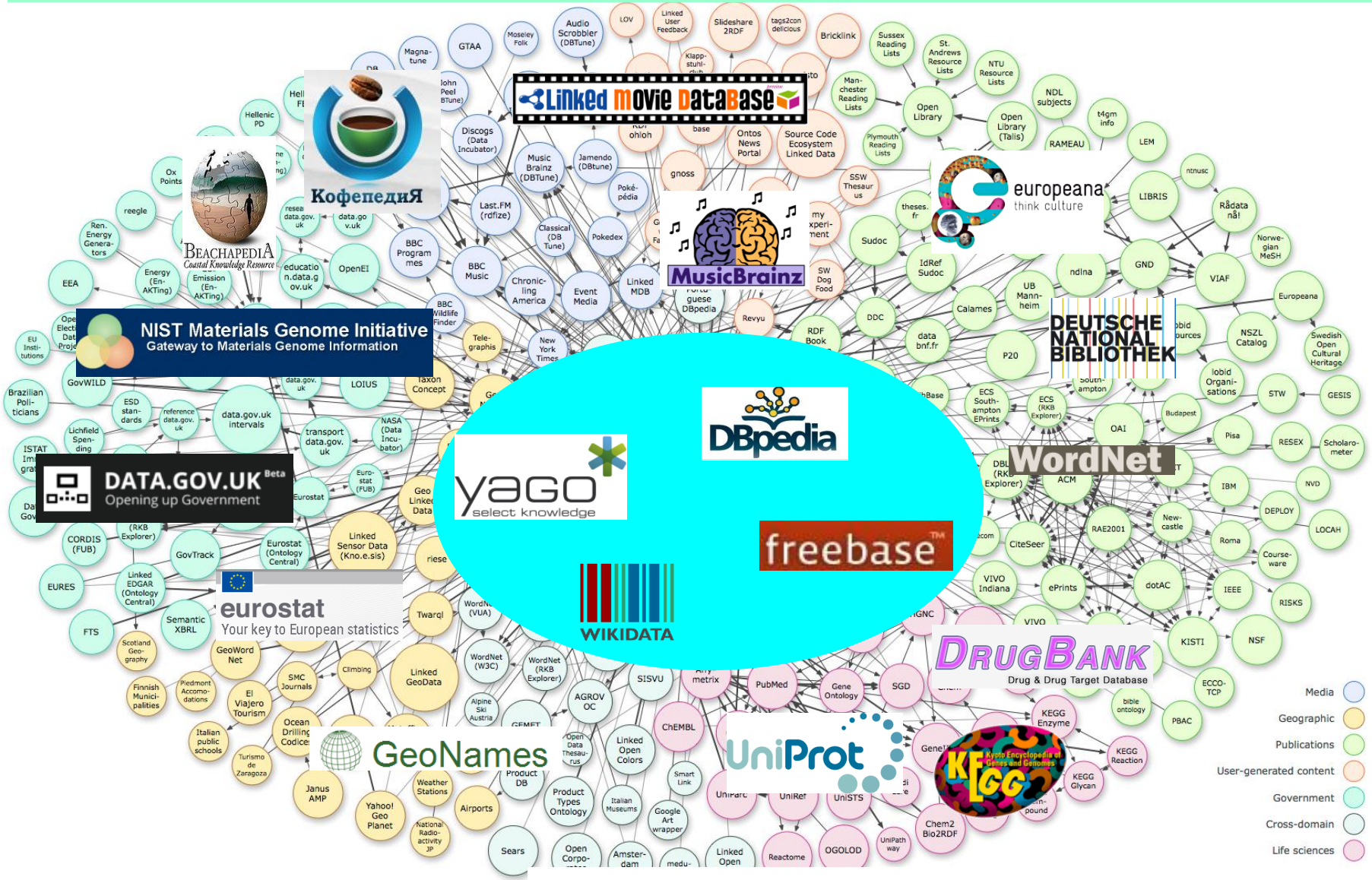
machine reading





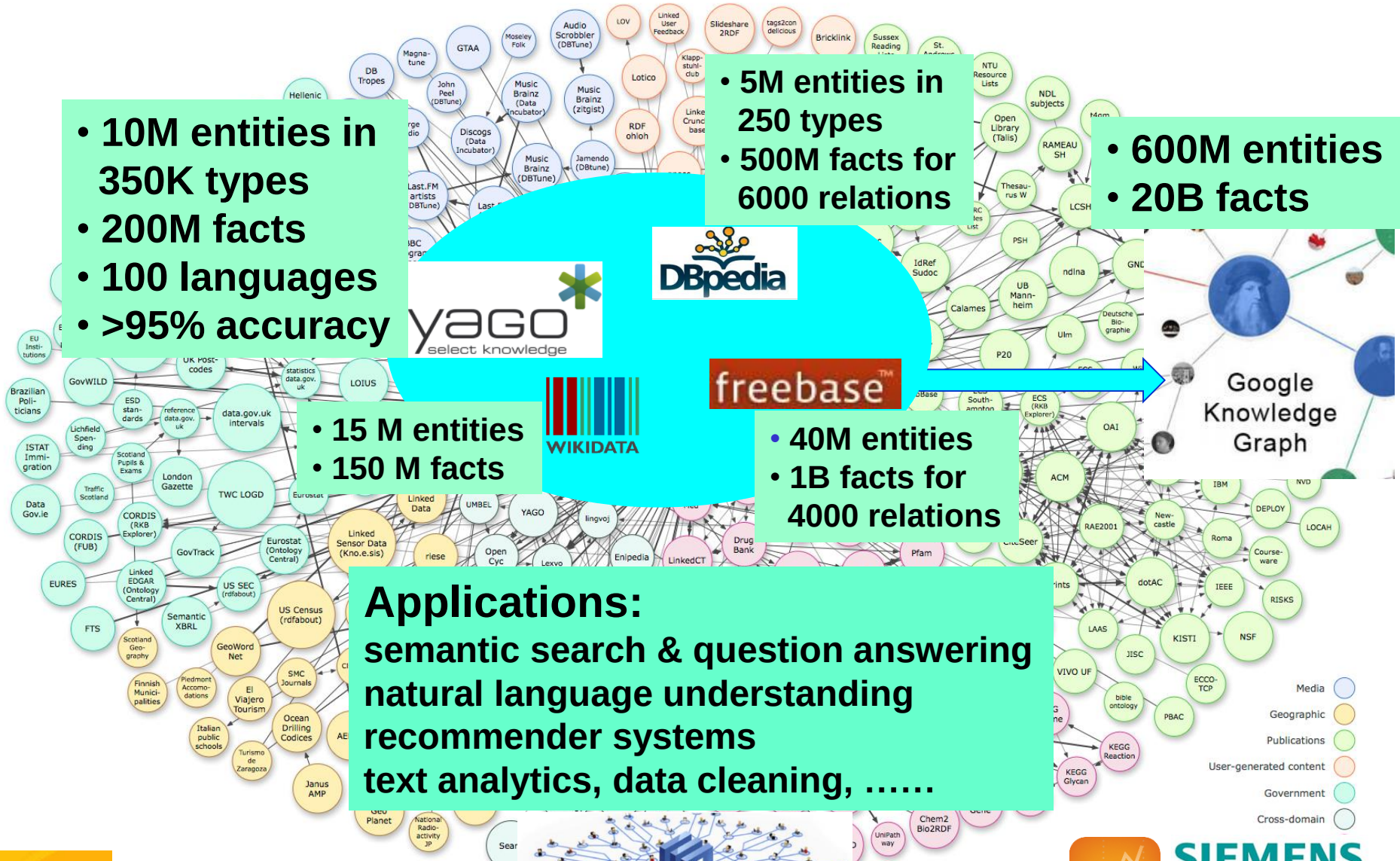
Web of Knowledge and Open Data

> 100 Billion subject-predicate-object facts from > 1000 sources



http://richard.cyganiak.de/2007/10/lod/lod-datasets_2011-09-19_colored.png

Web of Knowledge



bing WolframAlpha computational knowledge engine Walmart Bloomberg SIEMENS Ingenuity for life

Web of Knowledge

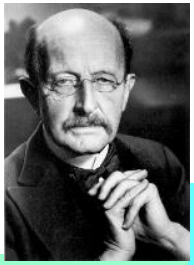
> 100 Billion subject-predicate-object facts from > 1000 sources

predicate (subject, object)

type (SteveJobs, entrepreneur) type (SteveJobs, computer architect) subtypeOf (entrepreneur, businessperson)	taxonomic knowledge
hasFounded (SteveJobs, Apple) hasDaughter (SteveJobs, LisaBrennan) namedAfter (AppleLisa, LisaBrennan) diedOf (SteveJobs, pancreatic cancer)	factual knowledge
hasSymptom (pancreatic cancer, jaundice) treats (ErlotinibHydrochloride, pancreatic cancer)	domain expert knowledge

Machine Knowledge for Answer Engines

Precise and concise answers for advanced information needs:



properties of entity

★ **Nobel laureate who outlived two world wars and all his children?**

sets of entities

★ **Politicians who are also scientists?**



relationships between entities

★ **Commonalities & relationships among:**
Kepler, Henri Poincaré, Liu Cixin, Zhang Jingchu?




Machine Knowledge for Answer Engines

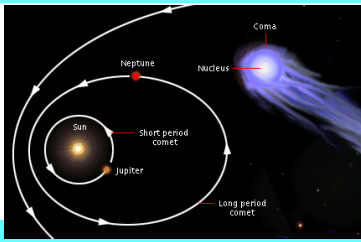
Precise and concise answers for advanced information needs:

properties of entity

★ **Nobel laureate who outlived two world wars and all his children?**

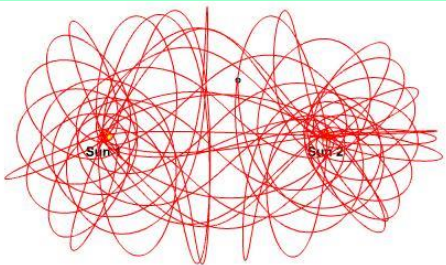
sets of entities

★ **Politicians who are also scientists?**



relationships between entities

★ **Commonalities & relationships among: Kepler, Henri Poincaré, Liu Cixin, Zhang Jingchu?**



Machine Knowledge for Answer Engines

Precise and concise answers
for advanced information needs:

real applications

- ★ **Proteins that bind to the Zika virus?**
- **Antidepressants that interfere with thyroid drugs?**
- **Polymer materials for super-capacitators?**
- **German philosophers influenced by William of Ockham?**
- **Books that influenced Liu Cixin?**
- **Green politicians mentioned in Panama Papers?**

Outline

- ✓ What Computers Know
- ★ What Computers Don‘t Know …
- ★ … and What Can Be Done About It
- ★ Conclusion

Missing on Predicates & Salient Facts

Which salient facts about an entity are in infoboxes?

WIKIPEDIA
The Free Encyclopedia

Liu Ci

Born	23 June Yangqua
Occupation	Science engineer
Nationality	Chinese
Period	1999–pr
Genre	Hard scie
Notable works	*Three-Body* trilogy

Baidu 百科

Chinese name	Liu Cixin
Foreign name	Liu, Cixin [10]
Alias	Liu electrician
Country of Cit...	China
place of birth	Beijing
date of birth	June 23, 1963
Occupation	Engineer, writer
graduated sc...	North China University of Water Res
	Hydropower
Major achiev...	From 1999 to 2006, he won consecu
	Fi Galaxy Award
Major achiev...	" Trisomy III" won the 2011 Global Ch
	Fiction Nebula Award for Best Novel
	trisomy III" won the " contemporary "
	of 2011 [4] " three-body " won the 20
	the West Lake · Type Literature bienn
	Gold [4] " trisomy III" won the ninth N
	excellent children's Literature Award
	categories) [4]

workedAs
(Liu Cixin, computer engineer)

translatedBy
(Liu Cixin, Liu Ken)

hasFavoriteBooks
(Liu Cixin,
{ Arthur C. Clarke: 2001,
George Orwell: 1984 })

locationInBook
(Three Body, Tsinghua Univ)

locationInBook
(Three Body, AlphaCentauri)

citesInBook
(Liu Cixin, Dark Forest,
Goethe:
“If I love you, what business is it of yours?“)

not in any KB !

Spectrum of Digital Knowledge (1): School Education for Computers

taxonomic knowledge:

type (SteveJobs, entrepreneur), subtypeOf (entrepreneur, businessperson)
type (SteveJobs, inventor), subtypeOf (inventor, human)
type (SteveJobs, YogaPractitioner), type (SteveJobs, GratefulDeadFan)

long-tail classes

factual knowledge:

hasFounded (SteveJobs, Apple), CEO (SteveJobs, Apple)
hasDaughter (SteveJobs, LisaBrennan), namedAfter (AppleLisa, LisaBrennan)
hasFavoriteSong (SteveJobs, Imagine), hasFavoriteSong (SteveJobs, Truckin‘)

long-tail entities

dated (SteveJobs, JoanBaez), admired (SteveJobs, BobDylan)
composed (JoanBaez, Diamonds&Rust), lyricsAbout (Diamonds&Rust, BobDylan)
sangAt (JoanBaez, memorialForSteveJobs)

long-tail relations

spatial & temporal knowledge:

diedOn (SteveJobs, 5-Oct-2011), diedIn (SteveJobs, Palo Alto)
happened (hasFounded (SteveJobs, Apple), Cupertino, 1976)
validDuring (CEO (SteveJobs,Apple), 1997-2011)

Spectrum of Digital Knowledge (2): Kindergarten and University

commonsense properties:

property (lemon, yellow), property (lemon, juicy), property (lemon, sour), ability (fish, swim), ability (human, speak), usedFor (classroom, teaching), maxHeight (human, 2.5 m), maxLength (snake, 10 m)

commonsense rules:

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

$\forall$ x, y, z mother (x,y) $\wedge$ mother (x,z) $\Rightarrow$ y = z

domain-specific expert knowledge:

type (Ubiquinone-8, coenzyme), expresses (COQ8, Ubiquinone-8) causes (lack of Ubiquinone-8, mitochondrial disorder)

Spectrum of Digital Knowledge (3): Learned in Life

socio-cultural and social knowledge:

invented (computer, Eckert and Mauchley, USA),
invented (computer, KonradZuse, Germany),
invented (computer, AlanTuring, UK), invented (computer, SteveJobs, young nerds)
drink (beer, Germany), drink (wine, California), drink (lassi, India)
alleviates (ice, bruises), alleviates (eucalyptusOil, sinusitis)

belief knowledge:

believe (Ptolemy, center (world,earth), believe (Galileo, center (world, sun))
believe (Chinese, badLuckNumber (4)), believe (Germans, badLuckNumber (13))
believe (AustralianAborigines, taboo (photosOfDeadPeople))

process knowledge:

type (changeTire, mechanicalTask)
subtask (changeTire, loosenBolts), subtask (changeTire, liftCar),
requires (loosenBolts, spiderWrench), requires (liftCar, jack)
precedes (loosenBolts, liftCar)

Knowledge Gaps

Temporal and Spatial Knowledge
Long-Tail Knowledge (on types and entities)
Dynamic Knowledge (events, emerging entities)
Open-Ended Knowledge (relation types)
On-the-Fly Knowledge
Visual Knowledge (on types and long-tail entities)
Cultural Knowledge
Commonsense Knowledge
Social Knowledge
Intensional Knowledge
Negative Knowledge

Outline

✓ **What Computers Know**

✓ **What Computers Don‘t Know …**

★ **… and What Can Be Done About It**

★ **Conclusion**

- **Open-Ended Knowledge**
- **Commonsense Knowledge**
- **Social Knowledge**

Open-Ended Relation Types


Ndapa Nakashole


Adam Grycner

Goal:
comprehensive repository of
binary predicates (and n-ary predicates)
with type signatures and paraphrases

Early work:
- **WordNet (Miller/Fellbaum), VerbNet (Palmer et al.)**
- **DIRT (Lin/Pantel: KDD‘01)**

Recent work:
- **PATTY (Nakashole et al.: EMNLP‘12)**
- **POLY (Grycner et al.: EMNLP‘16)**
- **Biperpedia (Gupta et al.: VLDB‘14)**
- **PPDB (Ganitkevich et al.: HLT-NAACL‘13)**
- **DEFIE (Bovi et al.: TACL‘15)**
- **FrameBase (Rouces et al.: ESWC‘15)**
- **schema.org**
- **more at Google, Microsoft, Baidu, … ?**

Paraphrases of Relations

Who performed or wrote music for which movie?

Morricone wrote the score for The Good, The Bad and The Ugly
Morricone's Ecstasy of Gold is in the soundtrack of the Good, the Bad, the Ugly
Beethoven's Elise is part of the score for Harry Potter 7
The soundtrack of The Fall includes the 2nd movement of the 7th, by Beethoven
Shakira gives her voice to Gazelle in the title song of Zootopia
The Zootopia trailer includes Shakira's title song
Zhang Ziyi's voice in the Beauty So appears in House of Flying Daggers
Andy Lau performs the title song of hat Women Want

score for: (Morricone, TheGoodTheBadT
(Beethoven, Harry Potter 7), …

soundtrack of: (Morricone, TheGoodTheBadT
(Beethoven, Harry Potter 7), …

voice in: (Shakira, Zootopia), (ZhangZiy

title song of: (Shakira, Zootopi (AndyLau, WhatWomenWant), …

- frequent sequence mining for relational phrases
- support sets of entity pairs for paraphrases
- clustering for "synsets"

musicInMovies (<musician>, <movie>):
voice in, title song of, soundtrack of, score for, song appears in, …

Semantically Typed Paraphrases of Relations

[Nakashole et al.: EMNLP'12, VLDB'12
Grycner et al.: EMNLP'15, EMNLP'16]

WordNet-style dictionary/taxonomy for relational phrases based on SOL patterns (syntactic-lexical-ontological)

Relational phrases are typed

<person> graduated from *<university>*
<singer> covered *<song>* *<book>* covered *<event>*

Relational phrases can be synonymous

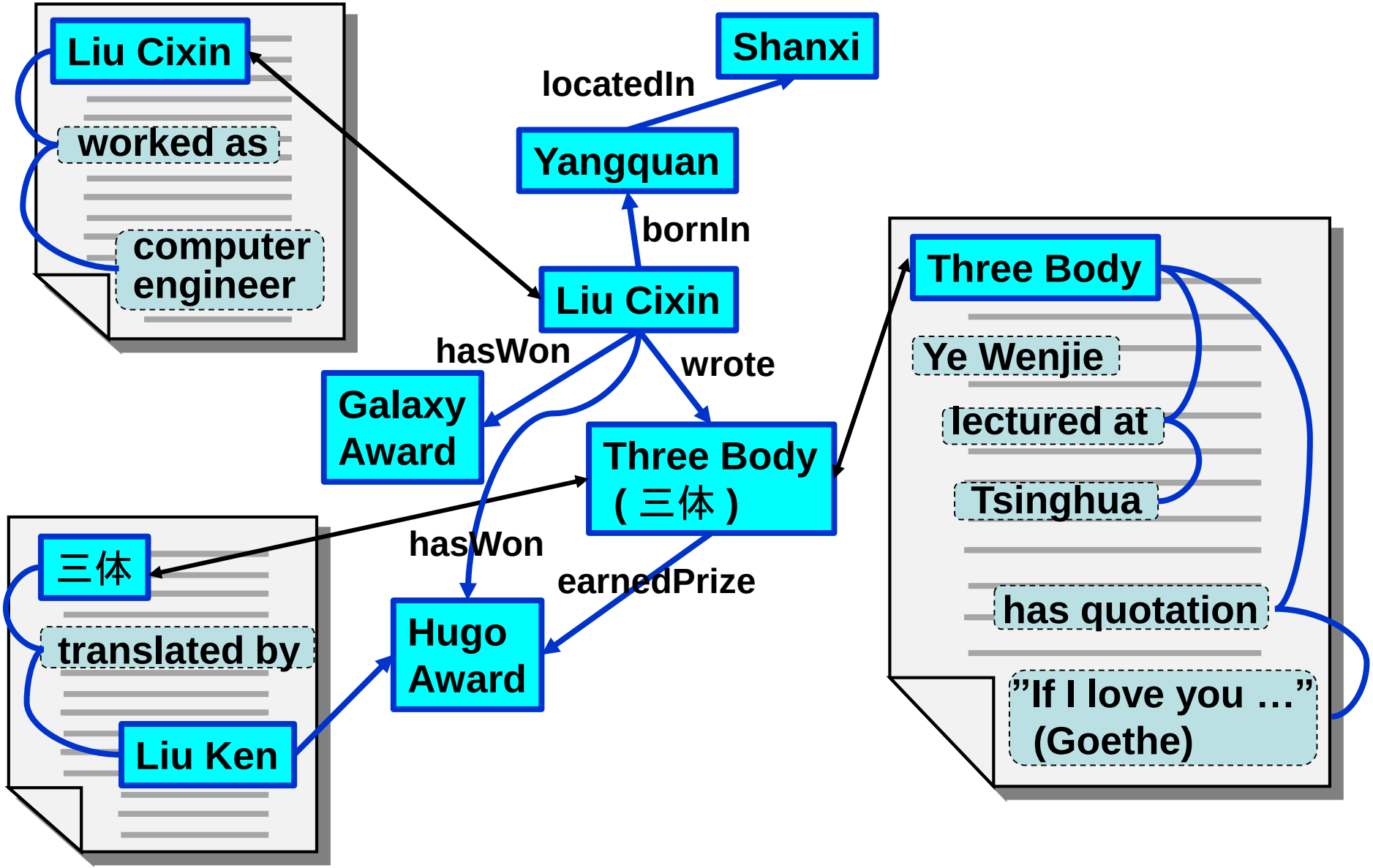
"graduated from" ⇔ "obtained degree in * from"
"and PRP ADJ advisor" ⇔ "under the supervision of"

One relational phrase can subsume another

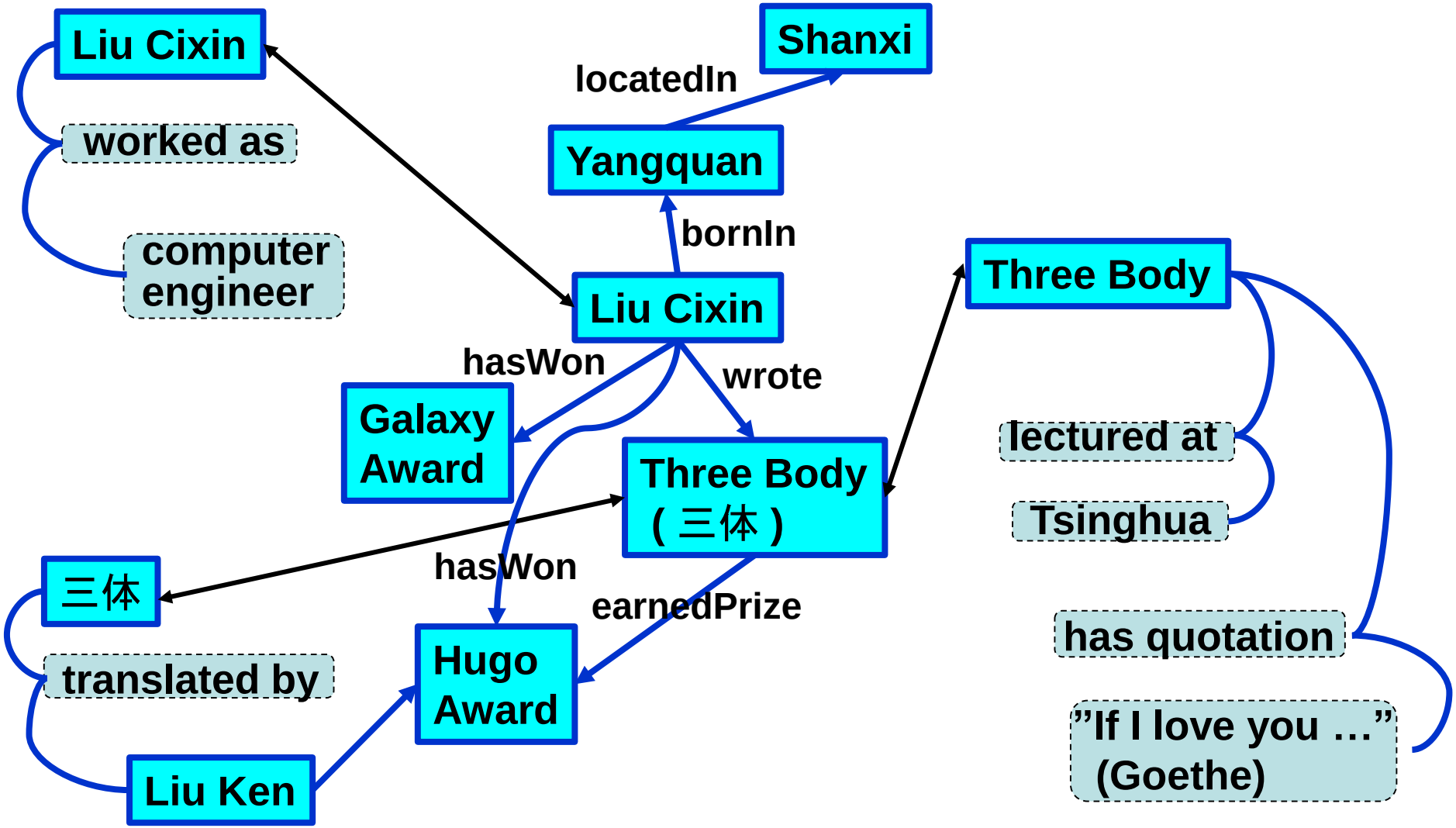
"wife of" ⇒ " spouse of"

350 000 SOL patterns with 4 Mio. instances
accessible at: www.mpi-inf.mpg.de/yago-naga/patty

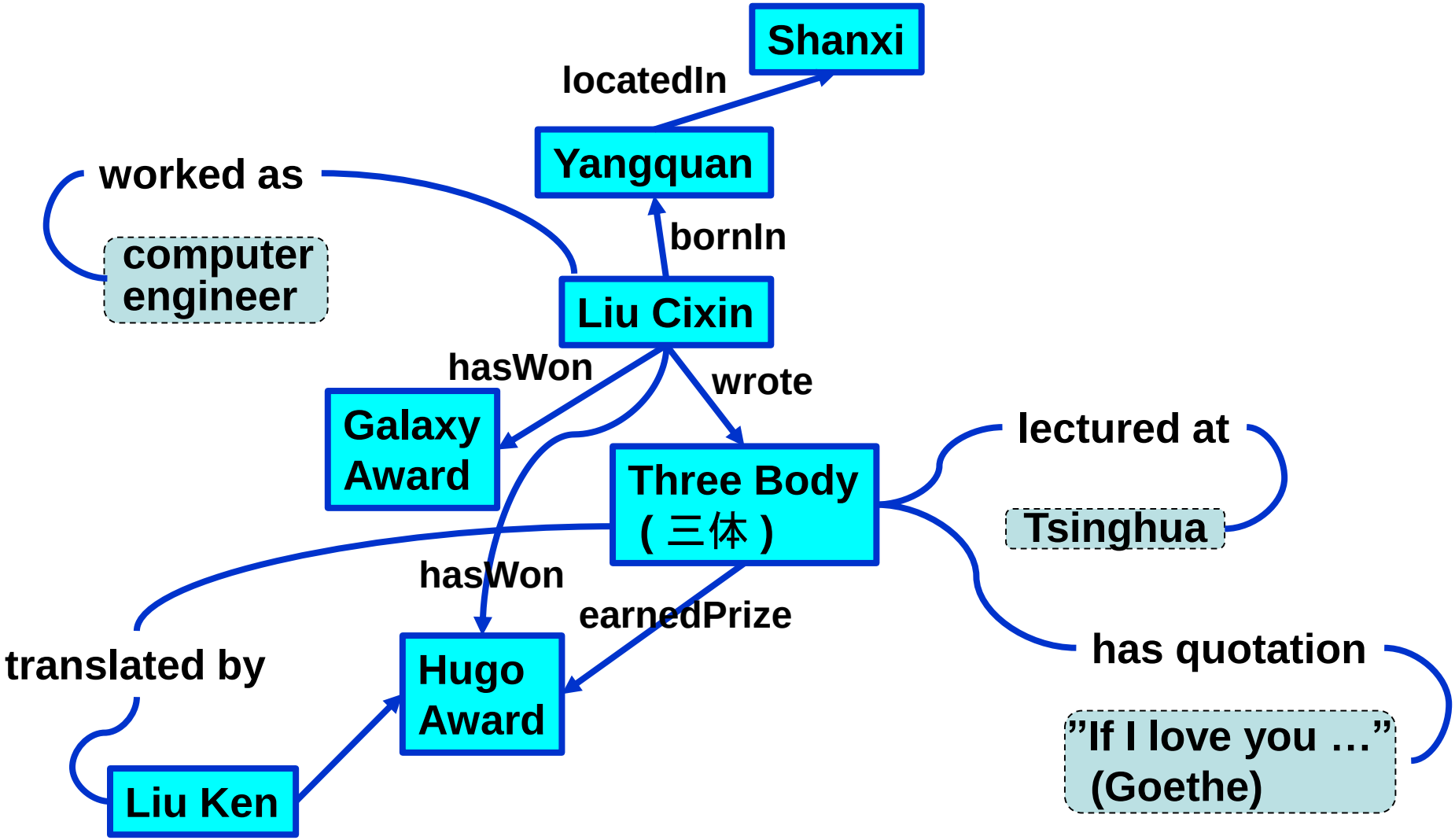
Extended Knowledge Graph (XKG): Connecting Facts with Text



Extended Knowledge Graph (XKG): Connecting Facts with Text



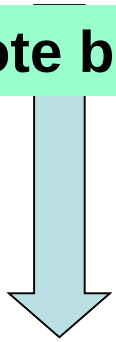
Extended Knowledge Graph (XKG): Connecting Facts with Text



Querying the XKG

(M. Yahya et al.: WSDM‘16, VLDB‘16)

engineers who wrote books that cite Goethe



automatically translate natural language question into structured query

```
Select ?x Where {
?x          “work“        “engineer“    .
?x          wrote         ?y            .
?y          {“quote“, “say“}  ?t        .
?t          {“work of“, “by“}  “Goethe“  . }
```

Triple patterns bind to S, P, O from KB or from text
Relaxations are generated automatically

Outline

✓ What Computers Know

✓ What Computers Don‘t Know …

★ … and What Can Be Done About It

★ Conclusion

- Open-Ended Knowledge
- Commonsense Knowledge
- Social Knowledge

Commonsense Knowledge: Not So Common

Every child knows that

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

pots and pans are in the kitchen or cupboard, on the stove, …
but not in in the bedroom, in your pocket, in the sky, …

children usually live with their parents

But: commonsense is rarely stated explicitly
Plus: web and social media have reporting bias

color of elephants ?
pink elephant: 0.9 Mio on Google
grey elephant: 0.4 Mio on Google

rich family: 27.8 Mio on Bing
poor family: 3.5 Mio on Bing

singers: 22.8 Mio on Bing
workers: 14.5 Mio on Bing

Pattern-based Harvesting of Commonsense Properties

Gerard de Melo

Niket Tandon

Approach: Start with seed facts

hasProperty (apple, round)
hasAbility (dog, bark)
hasLocation (plate, table)

Learn patterns that express these relations, such as

X is very Y, X can Y, X put in/on Y, …

Apply patterns to Web, books, N-grams corpora, image tags, etc.
→ statistics, semisupervised learning, constraint reasoning

hasColor (elephant, grey), hasShape (circle, round) …
hasAbility (fish,swim), hasAbility (human, talk) …
usedFor (book, learn), usedFor (computer, learn)
partOf (wheel, bike), partOf (wheel, car) …
hasTemperature (oven, hot), hasTaste (chili, hot) …

WebChild KB:
5 Mio. assertions
semantically typed
sense-disambiguated

Commonsense & Visual Contents

[N. Tandon et al.: WWW 15, CIKM 15, AAAI 16]

Refined part-whole relations from web&books text and image tags

→ 6.7 Mio sense-disambiguated triples for physicalPartOf, visiblePartOf, hasCardinality, memberOf, substanceOf

trafficJam:…

Activity knowledge from movie&TV scripts, aligned with visual scenes

→ 0.5 Mio activity types with attributes: location, time, participants, prev/next

www.mpi-inf.mpg.de/yago-naga/webchild/

Human Activities in Movie Scripts

STAR WARS: THE FORCE AWAKENS
Written By Lawrence Kasdan & J.J. Abrams and Michael Arndt

INT. MAZ' CASTLE

Rey steps down into the basement corridor.
BB-8 nervously follows her.
…
Rey OPENS THE BOX and sees inside
Luke Skywalker's original lightsaber.

activity: follow
agent: robot, AI
participant: woman
location: basement, indoor

activity: open box
agent: woman
participants: box, lightsaber
location: basement, indoor

Human Activities in Movie Scripts

STAR WARS: THE FORCE AWAKENS
Written By Lawrence Kasdan & J.J. Abrams and Michael Arndt

STARKILLER BASE - OSCILLATOR STRUCTURE – NIGHT

KYLO REN: I know what I have to do, but I don't know if I have the strength to do it. Will you help me?

HAN SOLO: Yes. Anything.

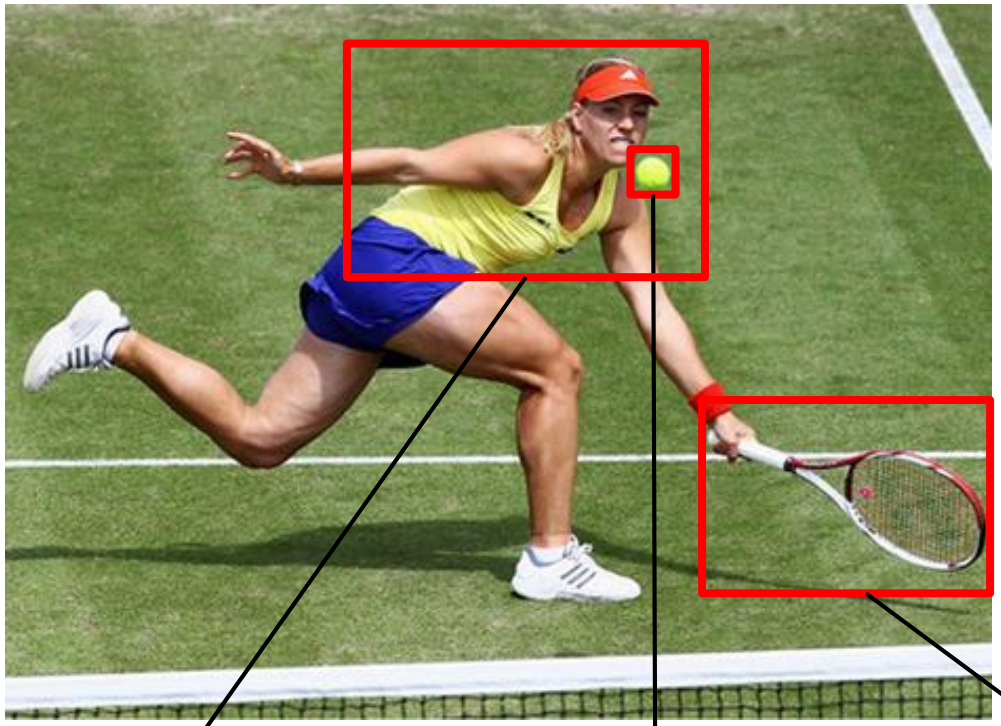
Kylo Ren unholsters his lightsaber and slowly extends it to Han. Han actually smiles and reaches out for the dark weapon. Kylo Ren ignites the lightsaber – THE FIERY BLADE SHOOTS SHOOTS RIGHT THROUGH HAN'S CHEST AND BACK!

activity: **shoot, kill**
agent: **son**
particpant: **father**
particpants: **lightsaber, weapon**

Commonsense for Language Understanding

- **How-to queries:**
  repair a bike tire, pitch a tent, cross a river, …
- **Scene search (over videos, books,):**
  romantic dinner, dramatic climb, …
- **Question disambiguation:**
  jordan weather forecast
  Germany‘s keeper at the Euro ?
- **Sentiment analysis:**
  the hottest place on earth
  the hottest bar in town

Commonsense for Computer Vision



woman (o1) **? lemon (o2)** **racket (o3)**

+ commonsense knowledge:

hasColor (tennisball, yellow)
hasShape (tennisball, round)
occursAt (tennisball, tennis court)
occursAt (racket, tennis court)

⇒ tennisball(o2)

Commonsense for Image Description

cake, 9 boys, 6 girls, 17 crowns
→ boys and girls with crowns around cake

+ commonsense knowledge: **→ children‘s birthday party**

Challenge: Commonsense Rules

Horn clauses:
can be learned by Inductive Logic Programming

$\forall$ x,m,c: type(x,child) $\wedge$ mother(x,m) $\wedge$ livesIn(m,t)) $\Rightarrow$ livesIn(x,t)
$\forall$ x,m,f: type(x,child) $\wedge$ mother(x,m) $\wedge$ spouse(m,f) $\Rightarrow$ father(x,f)

Advanced rules beyond Horn clauses:
specified by human experts

$\forall$ x: type(x,spider) $\Rightarrow$ numLegs(x)=8
$\forall$ x: type(x,animal) $\wedge$ hasLegs(x) $\Rightarrow$ even(numLegs(x))
$\forall$ x: human(x) $\Rightarrow$ ($\exists$ y: mother(x,y) $\wedge$ $\exists$ z: father(x,z))
$\forall$ x: human(x) $\Rightarrow$ (male(x) $\vee$ female(x))
$\forall$ x: human(x) $\wedge \neg$ adopted(x) $\Rightarrow$ numParents(x)=2
$\forall$ x: pope(x) $\Rightarrow \neg$ ($\exists$ y: father(x,y))

Outline

✓ What Computers Know

✓ What Computers Don‘t Know …

★ … and What Can Be Done About It

★ Conclusion

- Open-Ended Knowledge
- Commonsense Knowledge
- Social Knowledge

Social Knowledge: “Folk Wisdom“

Beliefs become like facts
in specific socio-cultural contexts

invented (computer, Eckert and Mauchley, USA)
invented (computer, KonradZuse, Germany)
invented (computer, AlanTuring, UK)
invented (computer, SteveJobs, young nerds)
invented (Internet, Cerf and Kahn, computer scientists)
invented (Internet, Berners-Lee, laymen)
invented (Internet, Al Gore, Al Gore supporters)

Need to tap riskier sources (social media posts etc.)

Knowledge distilled from subjective behavior / experience of many people

- common habits on food, drinks, social behavior
- experience with electric/diesel/green cars, on the road
- grassroots insight on health, symptoms, therapies, …

Credibility and Trust: Where the Truth Lies




Subhabrata Mukherjee Cristian Danescu Kashyap Popat Jannik Strötgen

Assess **credibility** of statements / claims on the Internet and the **trustworthiness** of their sources

- **Search results:** ***love affairs of Hillary Clinton ?***
- **Biased news:** ***Merkel hates Greece***
- **KB contents:** ***Berners-Lee & Al Gore invented the Internet***
- **Social media:** ***Obamacare requires microchip implant***
- **Health forums:** ***Xanax causes hallucinations***

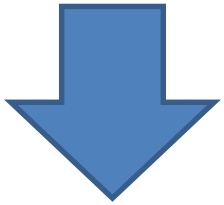
Crucial for info extraction, KB curation, explanation, opinion mining, web contents analytics, etc.

Credibility & Trust in Health Communities

I took the whole med cocktail at once. Xanax gave me wild hallucinations and a demonic feel.

Xanax and Prozac are known to cause drowsiness.

Xanax made me dizzy and sleepless.



hasSideEffects (Xanax, hallucinations)
hasSideEffects (Xanax, dizziness)
hasSideEffects (Xanax, insomnia)

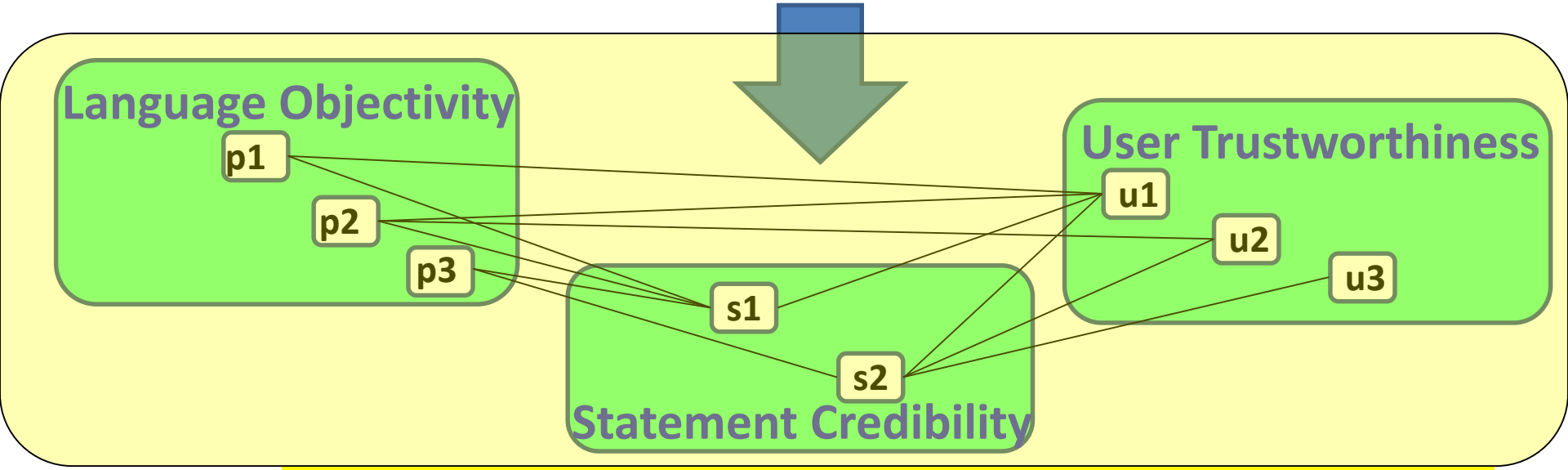
Credibility & Trust in Health Communities

[KDD 2014]

I took the whole med cocktail at once. Xanax gave me wild hallucinations and a demonic feel.

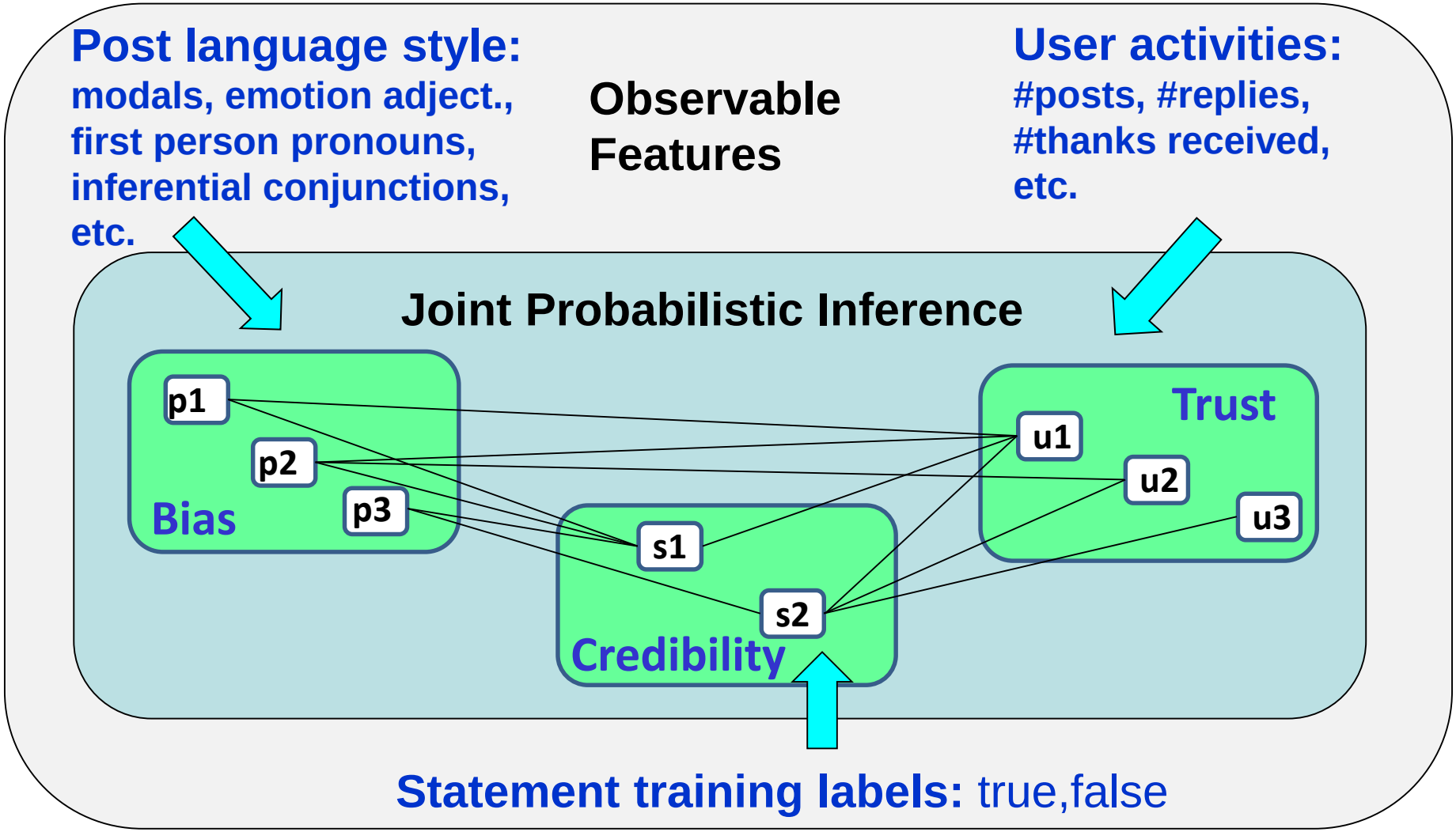
Xanax and Prozac are known to cause drowsiness.

Xanax made me dizzy and sleepless.



joint reasoning with probabilistic graphical model
(semi-supervised heterogeneous CRF with EM-style learning)

Semi-Supervised Heterogeneous MRF for Joint Inference on Posts, Users, Statements



Credibility of Textual Claims

Jack Ma is the new owner of AC Milan

[K. Popat et al.: CIKM 2016]

True or False ?

search Web for evidence

GAME ON

Chinese investors want to buy the legendary AC Milan soccer club from Silvio Berlusconi

By Zheping Huang | May 10, 2016

TIME

SIGN IN | SUBSCRIBE

LATEST | MAGAZINE | VIDEOS

Legendary Italian Soccer Club AC Milan Is Now Chinese-Owned

SPORTS SOCCER

Legendary Italian Soccer Club AC Milan Is Now Chinese-Owned

Charlie Campbell / Beijing @charliecamp6ell

July 6, 2016

"Milan has n

Storied Italian
been sold to a
the club's own
Minister Silvi

CNBC

HOME U.S. | NEWS | MARKETS | INVESTING | TECH | MAKE IT | VIDEO

FINANCE | BANKS | INVESTING | WALL STREET | HEDGE FUNDS | M&A | INSURANCE

Alibaba founder Jack Ma laughs off AC Milan football club purchase rumors

CNBC.com staff | @CNBC

Thursday, 28 Apr 2016 | 9:17 PM ET

CNBC

calciomercato.com

ITALY'S FIRST FOOTBALL NEWS SITE

Italian News | Like

CM EXCLUSIVE: Alibaba and Whampoa not interested in buying AC Milan. Is Robin LI the right man?

28 April at 12:50

Throug *Weibo*, the Chinese *Twitter*, **Jack Ma** dismissed the rumors regarding the possibility to buy **AC Milan's** shares.

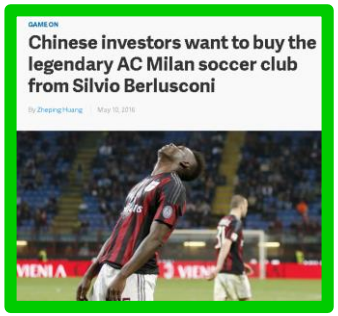
Last night, some sources reported of the interest of **Li Ka Shing** –owner of the **Hutchinson Whampoa Limited**- in buying the Lombard club's stake, but Li Ka –contacted by *calciomercato.com*- has denied such interest.

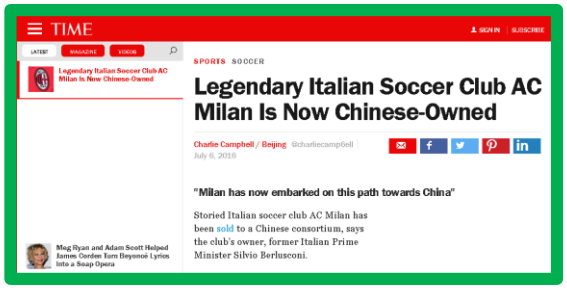
Credibility of Textual Claims

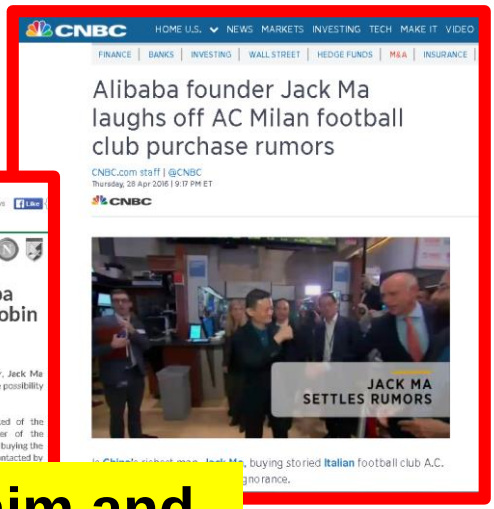
Jack Ma is the new owner of AC Milan

[K. Popat et al.: CIKM 2016]

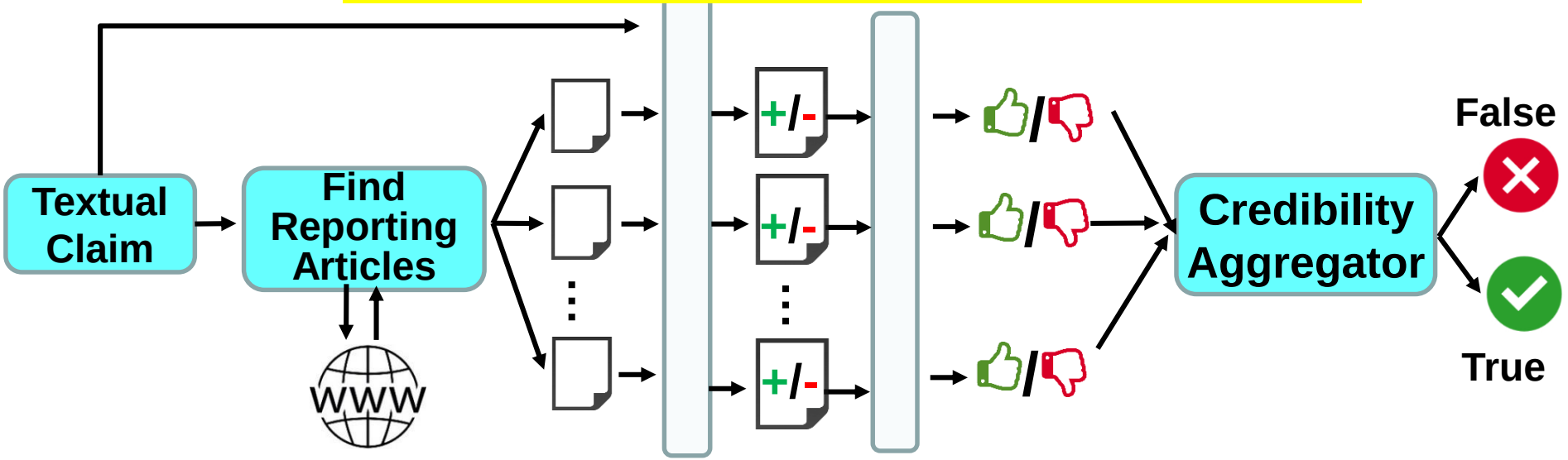
True or False !







joint reasoning on credibility of claim and authority of supporting/refuting sources



Research Opportunities:

- **Understand & incorporate the influence of attribution in sources and dissemination across sources**
- **Consider anonymous sources (e.g. user posts in online communities)**
- **Explain veracity of claims in socio-cultural context**

Outline

- ✓ What Computers Know
- ✓ What Computers Don‘t Know …
- ✓ … and What Can Be Done About It
- ★ Conclusion

What Computers Should Know: Research Opportunities

High-Coverage KB and Open-Ended KB Growth
need Relational Paraphrases
→ more predicates, more facts, XKG

Commonsense Knowledge crucial for
Language Understanding & Visual / Multimodal Contents
→ tap into images, videos, scripts, books

Socio-Cultural Knowledge
needs to tap into risky sources
→ analyze, assess, explain credibility & trust

Take-Home Message

Computers shouldn‘t believe everything

more knowledge, analytics, insight

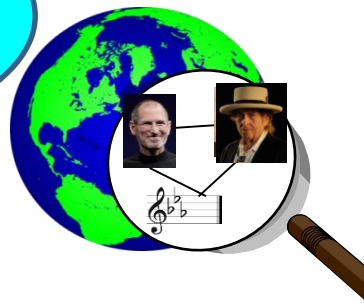
Computers need commonsense

knowledge acquisition

Internet Contents

Digital Knowledge

machine reading



Computers need more knowledge and socio-cultural knowledge