WebChild: Harvesting and Organizing Commonsense Knowledge from Web

Niket Tandon

Max Planck Institute for Informatics

Saarbrücken, Germany

Joint work with: Gerard de Melo, Fabian Suchanek, Gerhard Weikum

Why Computers Need Commonsense Knowledge

Who looks hot ?

pop-singer-n^1
hasAppearance
hot-a^3

What tastes hot ?

chili-n^1
hasTaste
hot-a^9

What is hot ?

volcano-n^1
hasTemperature
hot-a^1

Why Knowledge Bases Are Not Sufficient

Freebase
(+ Dbpedia,
 Yago, …)

Jay-Z bornOn 4-Dec-1969
Jay-Z bornIn   Brooklyn
Brooklyn locatedIn NewYorkCity
Jay-Z married Beyonce
…..

only facts about named entities

ConceptNet
(+ …)

pop-singer isa musician
pop-singer hasProperty hot
volcano hasProperty hot
action hasProperty hot
…..

hot ≠ hot ≠ hot

only hasProperty or relatedTo

Key Novelties of WebChild

1. **Fine-grained relations** for commonsense knowledge
   (derived from WordNet):
   hasAppearance, hasTaste, hasTemperature, hasShape,
   evokesEmotion, …..

2. **Sense-disambiguated arguments** of knowledge triples
   (mapped to WordNet):
   pop-singer-n^1 hasAppearance hot-a^3
   chili-n^2 hasTaste hot-a^9
   volcano-n^1 hasTemperature hot-a^1

Semantically refined commonsense triples

1. Extract generic: salsa   hasProperty   hot

Patterns

beautiful rose

salsa was really hot

...

<adj>   <noun>
<noun>   linking_verb   [adverb]   <adj>

Semantically refined commonsense triples

1. Extract generic: salsa   hasProperty   hot

2. Refine: salsa-n^1 hasTaste   hot-a^9

WordNet "salsa"

19 fine-grained relations

1. hasEmotion
2. hasSound
3. **hasTaste**
4. hasAppearance
…

WordNet "hot"

Semantically refined commonsense triples

Domain Population	Computing Assertion	Range Population

Refine: **salsa-n^1**   **hasTaste**   **hot-a^9**

what has taste   disambiguate, **classify**, rank   how does it taste

pizza-n^1
sauce-n^1
java-n^2
…

chocolate-n^2 , sweet-a^1
milk-n^1, tasty-a^1
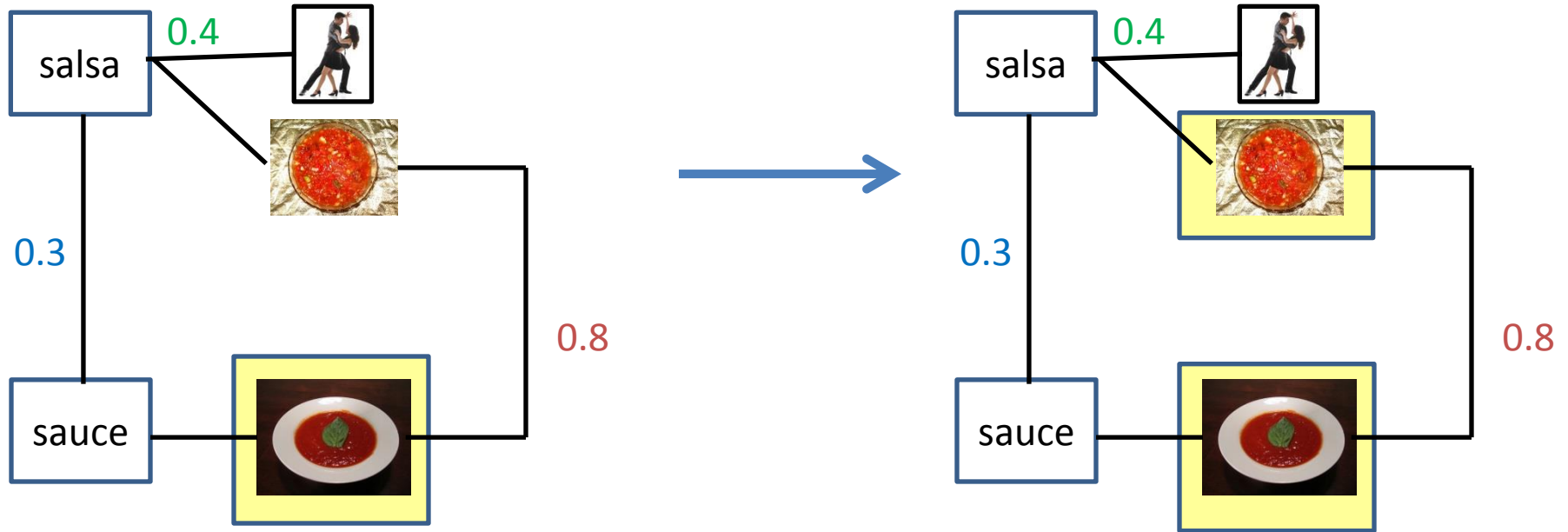…

spicy-a^1
hot-a^9
sweet-a^1
…

Graph construction per relation (e.g. hasTaste)

- Edge weight:

  taxonomic (between senses) ,

  co-occurrence statistics (between words),

  distributional (between word, senses).
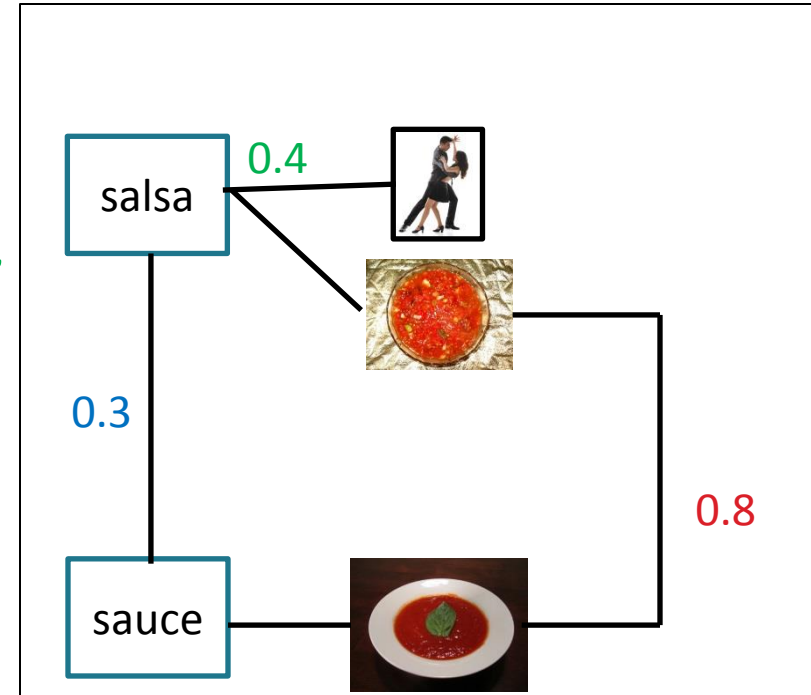
Label Propagation on constructed graph
for domain of hasTaste

$$\left[\mu_1 \sum_v p_v^{inj} \sum_l (Y_{vl} - \hat{Y}_{vl})^2 + \mu_2 \sum_{v,u} p_v^{cont} W_{vu} \sum_l (Y_{vl} - \hat{Y}_{ul})^2 + \mu_3 \sum_{vl} (\hat{Y}_{vl} - R_{vl})^2 \right]$$

WebChild : Model

Domain (hasTaste)   **Range (hasTaste)**   **Assertions (hasTaste)**
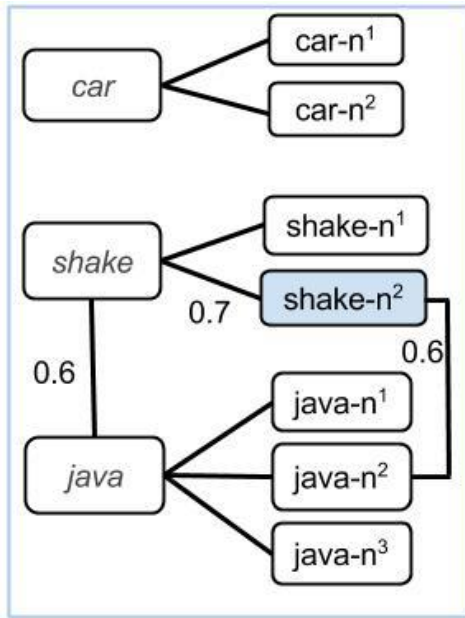
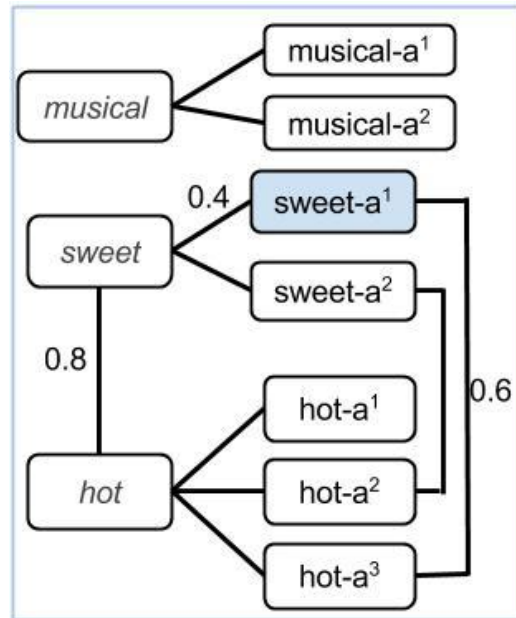Seed label loss **+** Similar node diff label loss **+** Regularize

$$\left[\mu_1 \sum_v p_v^{inj} \sum_l (Y_{vl} - \hat{Y}_{vl})^2 + \mu_2 \sum_{v,u} p_v^{cont} W_{vu} \sum_l (Y_{vl} - \hat{Y}_{ul})^2 + \mu_3 \sum_{vl} (\hat{Y}_{vl} - R_{vl})^2 \right]$$
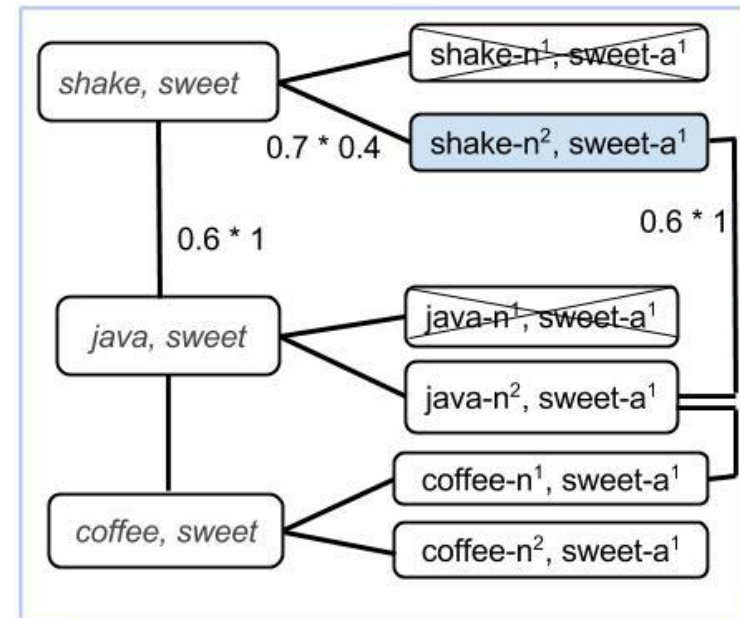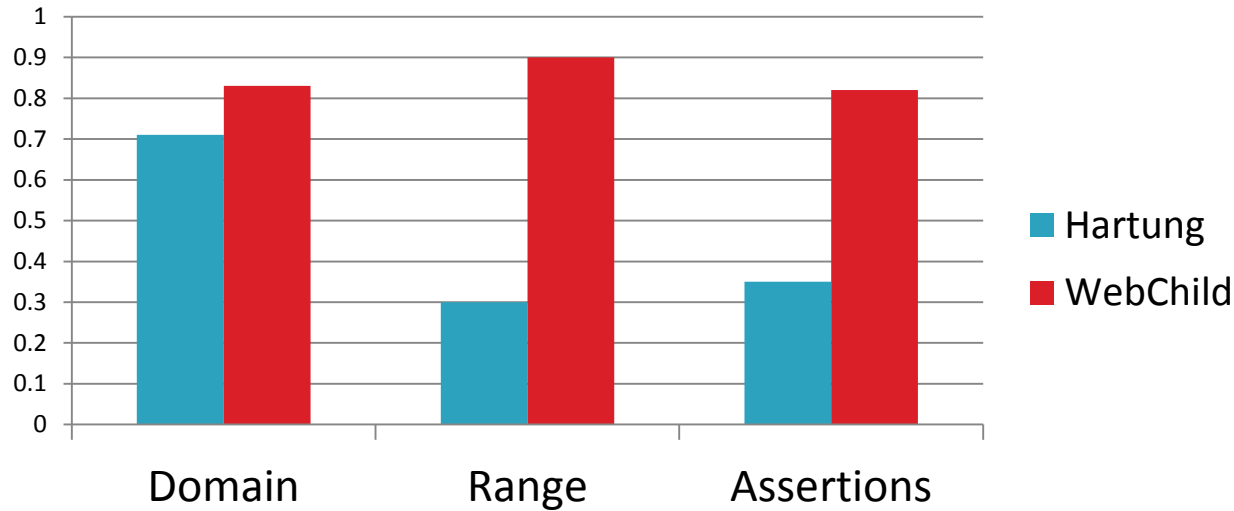
Experiments

Accuracy: over manually sampled data.

Statistics: Large, semantically refined commonsense knowledge.

	#instances	Precision
Noun senses	221 K	0.80
Adj senses	7.7 K	0.90
Assertions	4.6 M	0.82

WebChild: Examples

Domain (hasShape)	Range (hasShape)	Assertions (hasSshape)
face-n^1	triangular-a^1	lens-n^1, spherical-a^2
leaf-n^1	tapered-a^1	palace-n^2, domed-a^1
...	...	...

Set expansion for: keyboard-n^1

Top 10 adjectives	ergonomic, foldable, sensitive, black, comfortable, compact, lightweight, comfy, pro, waterproof
Top 5 expansions	keyboard, usb keyboard, computer keyboard, qwerty keyboard, optical mouse, touch screen

Set expansion for: keyboard-n^2

Top 10 adjectives	universal, magnetic, small, ornamental, decorative, solid, heavy, white, light, cosmetic
Top 5 expansions	wall mount, mounting bracket, wooden frame, carry case, pouch

Conclusion

- Graph methods help overcome sparsity of commonsense in text.

- **WebChild**: First commonsense KB with fine-grained relations and disambiguated arguments ; 4.6 million assertions including domain and range for 19 relations. Publically available at:

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

Additional slides.

Use Case: Set Expansion

Output: top ranked adjectives and similar nouns (cosine over attributes) .

Input: chocolate-n^2

Top 10 adjectives	smooth, assorted, dark, fine, delectable, black, decadent, white, yummy, creamy
Top 5 expansions	chocolate bar, chocolate cake, milk chocolate, chocolate chip, chocolate fudge

Input: keyboard-n^1

Top 10 adjectives	ergonomic, foldable, sensitive, black, comfortable, compact, lightweight, comfy, pro, waterproof
Top 5 expansions	keyboard, usb keyboard, computer keyboard, qwerty keyboard, optical mouse, touch screen

Approach

For **range and domain population**:

Extract a large list of ambiguous (potentially noisy) candidates.

Construct a weighted graph of ambiguous words and their senses.

Mark few seed nodes in the graph.

Use propagation concept: similar nodes (beautiful) (lovely) have similar labels

For **computing assertion**:

Use the range and domain to prune search space of assertions (for a relation)

Use propagation concept: similar nodes (car, sweet) (car, lovely) similar labels.

Approach: Extract and refine

Google n-grams

Y/adj X/noun

→

X/noun
linking_verb
adverb Y/adj

red rose

rose was very beautiful

temperature was hot

Goal: Semantically refined commonsense properties
Connect nouns with adjectives via fine-grained relations
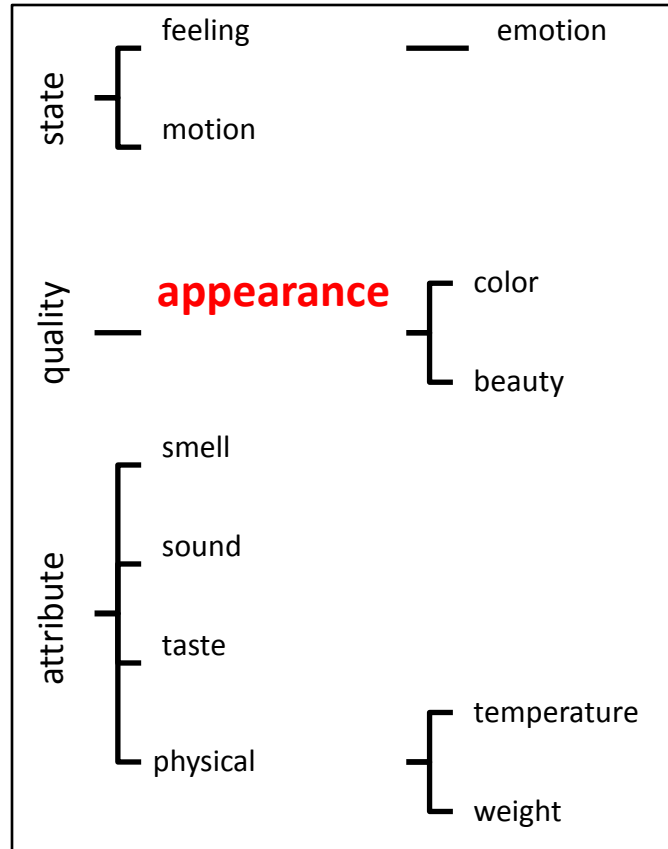
1. Extract: suit hasProperty hot

2. Refine: suit-n^2 quality . appearance hot-a^3

WordNet "suit"

1. Lawsuit
2. **Dress**
3. Playing card suit
4. …

state
— feeling
— motion

— emotion

quality
— **appearance**
— color
— beauty

attribute
— smell
— sound
— taste
— physical
— temperature
— weight

WordNet "hot"

1. Burning
2. Violent
3. **Stylish**
4. …

Experiments

Accuracy and coverage : manually sampled data.

System	Domain	Range	Assertions
Controlled LDA MFS (Hartung et al. 2011)	0.71	0.30	0.35
WebChild	0.83	0.90	0.82

Statistics: Large, semantically refined commonsense knowledge.

	#instances	Precision
Noun senses	221 K	0.80
Adj senses	7.7 K	0.90
Assertions	4.6 M	0.82

Related Work

	Commonsense Knowledge	Automatically constructed	**Unambiguous arguments**	**Fine-grained relations**
Linked Data	✖	✓	✓	✖
Cyc	✓	✖	✖	✖
Concept Net	✓	✓	✖	✖
WebChild	✓	✓	✔	✔

Goal: Semantically refined commonsense properties

1. Extract: mole   hasProperty   hot

2. Refine: mole-n^3 taste   hot-a^4

WordNet "mole"	19 fine-grained relations	WordNet "hot"
1. Gram molecule 2. Skin mark **3. Sauce** 4. Animal …	1. Emotion 2. Sound **3. Taste** 4. Appearance …	1. Burning 2. Violent 3. Stylish **4. Spicy** …

Goal: Semantically refined commonsense properties

Domain Population

Computing Assertion

Range Population

Refine: **mole-n^3**
in **domain** of taste

taste
disambiguate, **classify**, rank

hot-a^4
in **range** of taste

domain (taste)

pizza-n^1
sauce-n^1
java-n^2
…

assertion (taste)

salsa-n^1 , hot-a^4
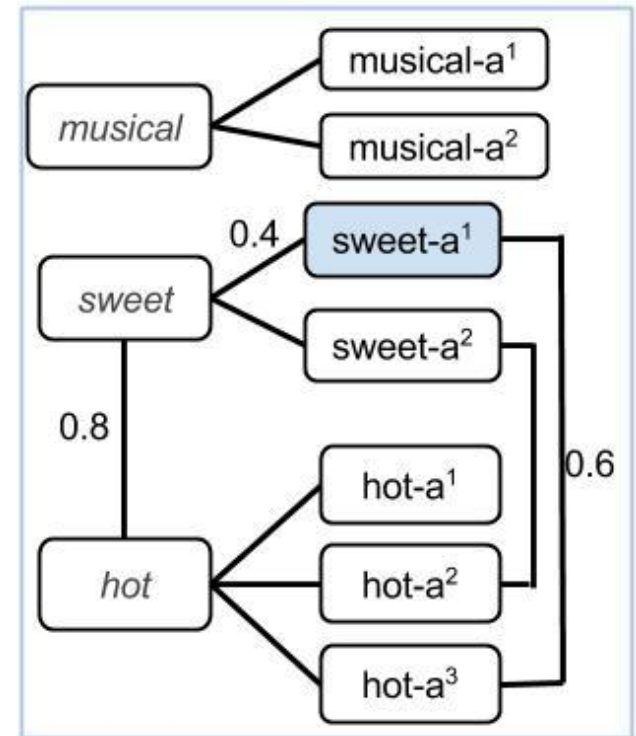chocolate-n^2 , sweet-a^1
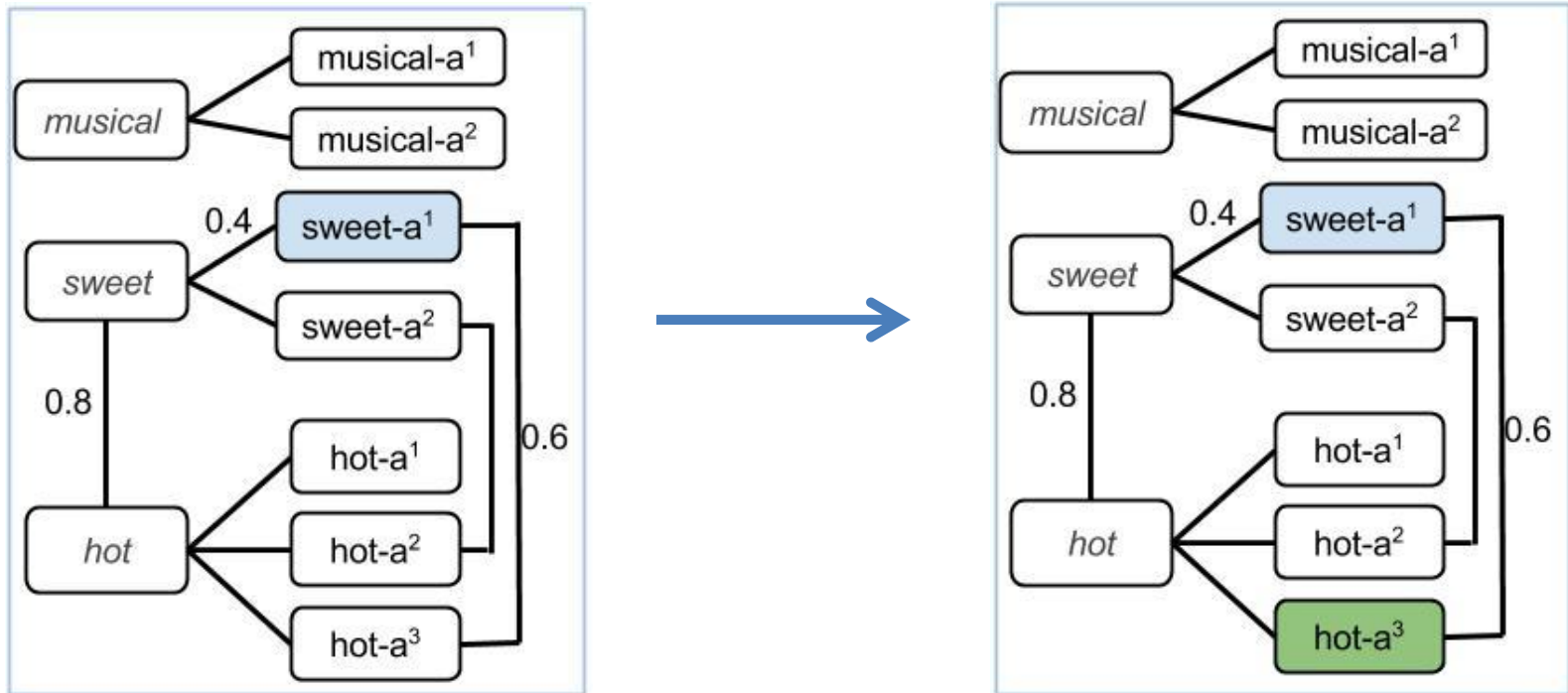milk-n^1, tasty-a^1
…

range (taste)

spicy-a^1
hot-a^4
sweet-a^1
…

Graph construction

- Edge weight:

  taxonomic (between senses) ,

  co-occurrence statistics (between words),

  distributional (between word, senses).

- One graph per attr. (here, **hasTaste**)

Label Propagation on constructed graph

$$\left[\mu_1 \sum_v p_v^{inj} \sum_l (Y_{vl} - \hat{Y}_{vl})^2 + \mu_2 \sum_{v,u} p_v^{cont} W_{vu} \sum_l (Y_{vl} - \hat{Y}_{ul})^2 + \mu_3 \sum_{vl} (\hat{Y}_{vl} - R_{vl})^2 \right]$$

WebChild: Examples

	Domain	Range	Assertions
hasTaste	strawberry-n^1	sweet-a^1	biscuit-n^2, sweet-a^1
	java-n^2	hot-a^9	chilli-n^1, hot-a^9
hasShape	face-n^1	triangular-a^1	lens-n^1, spherical-a^2
	leaf-n^1	tapered-a^1	table-n^2, domed-a^1

Set expansion for: keyboard-n^1

Top 10 adjectives	ergonomic, foldable, sensitive, black, comfortable, compact, lightweight, comfy, pro, waterproof
Top 5 expansions	keyboard, usb keyboard, computer keyboard, qwerty keyboard, optical mouse, touch screen

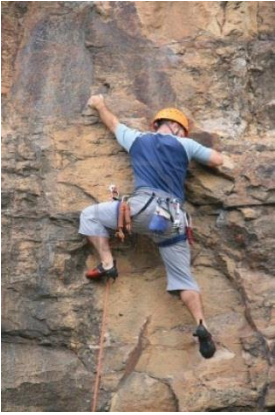# Why Computers Need Commonsense Knowledge

Who looks cool ?

Who lives cool ?

Commonsense Knowledge

- Image search query: "adventurous person" should also match an image of a man "**rock climbing**" (evokes emotion "thrilling")

- What is **red**, **edible**, **tasty** and **soft**?

- What is **similar** to chocolate bar, but **soft**?
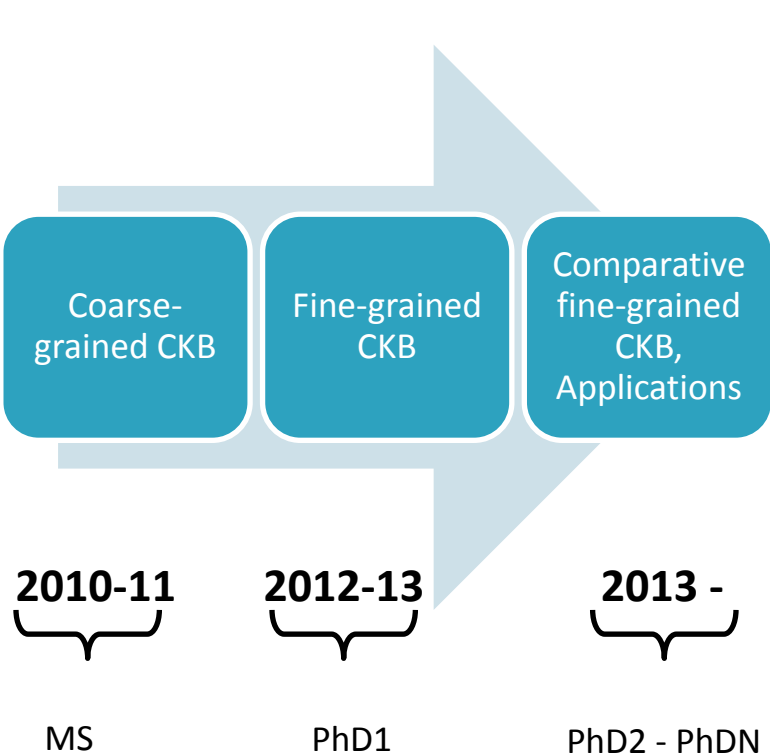
Why Computers Need Commonsense Knowledge
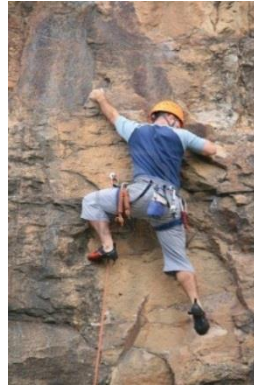
Who looks cool ?

Who lives cool ?

Commonsense from the Web

Coarse-grained CKB → Fine-grained CKB → Comparative fine-grained CKB, Applications

2010-11 **2012-13** **2013 -**

MS PhD1 PhD2 - PhDN

Niket Tandon

Supervisor: Prof. Gerhard Weikum

Collaborator: Prof. Gerard de Melo

Max Planck Institute for Informatics

- Image search query: "adventurous person" should also match an image of a man "**rock climbing**" (evokes emotion "thrilling")

- What is **red**, **edible**, **tasty** and **soft**?

What is **similar** to chocolate bar, but **soft**?

⚠ No results found for **"cherries are sweeter than oranges"**.

Commonsense from the Web

	Commonsense Knowledge	Automatically constructed	**Unambiguous arguments**	**Fine-grained relations**
Linked Data	✖	✔	✔	✖
Cyc	✔	✖	✖	✖
Concept Net, Tandon AAAI'11	✔	✔	✖	✖
WebChild WSDM'14	✔	✔	✔	✔