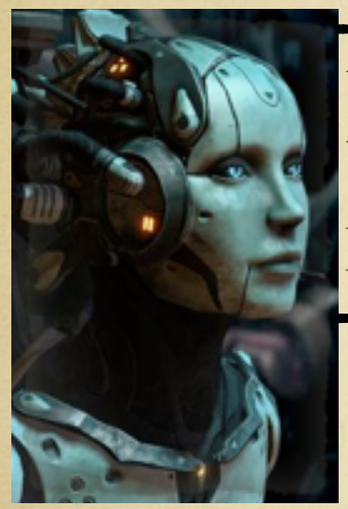
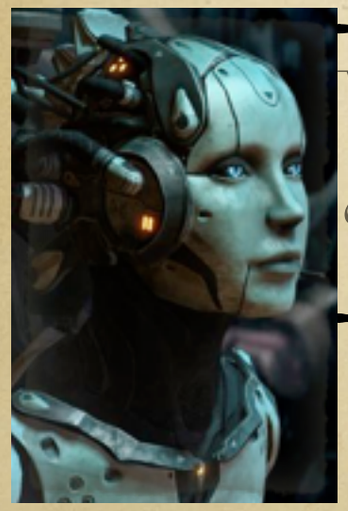
Knowledge Base

Robot in a room…



I can recognize everything in the room (proudly)

Bring me a cup of hot water



Well, I can tell you "where is the cup?"

- Recognize everything, but can do nothing

What is missing?

Bring me a cup of hot water

- find a cup
- realize a cup has *containable* affordance

Affordance

- A cup
  - grasp
  - filled in water
  - pour

Attribute

- A cup
  - brittle
  - made of glass, plastic
  - has a handle

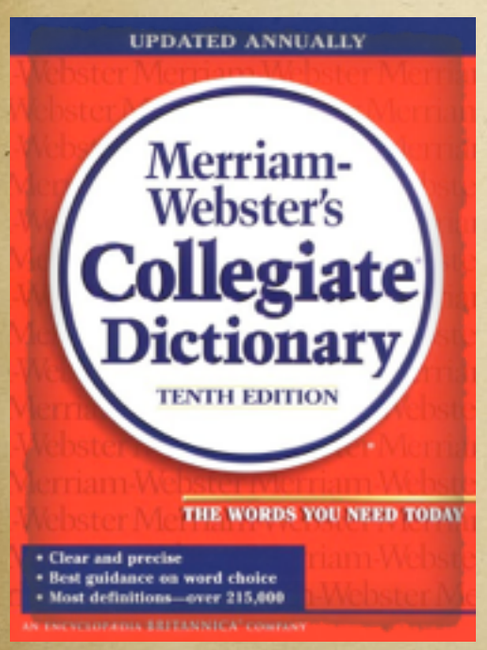
What is missing?

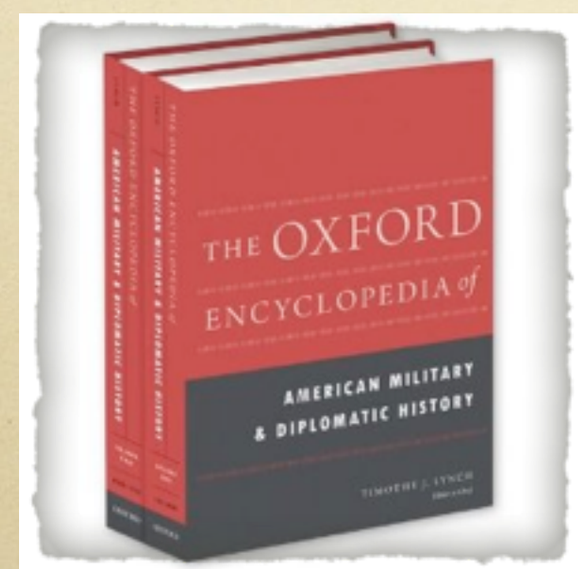
Bring me a cup of hot water

- find a cup
- realize a cup has *containable* affordance
- cup is empty
- find tape, fill in water
- find microwave
- heat it up

The Common Knowledge

The Common Knowledge

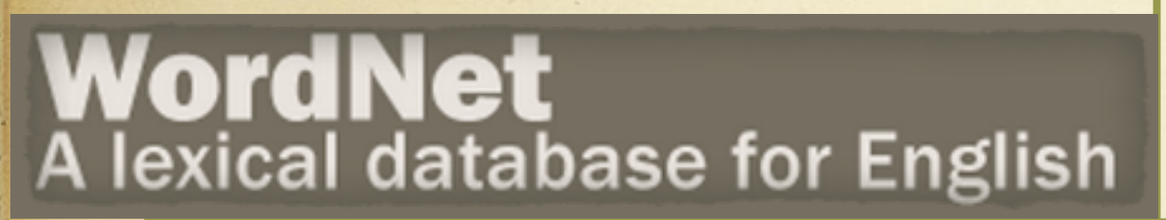




IMAGENET

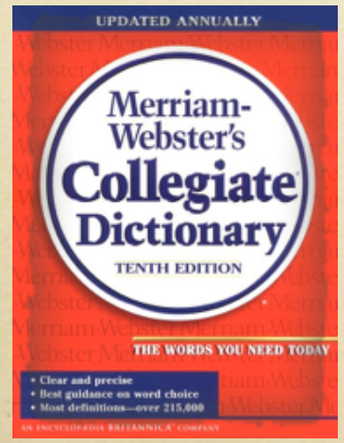
WordNet
A lexical database for English

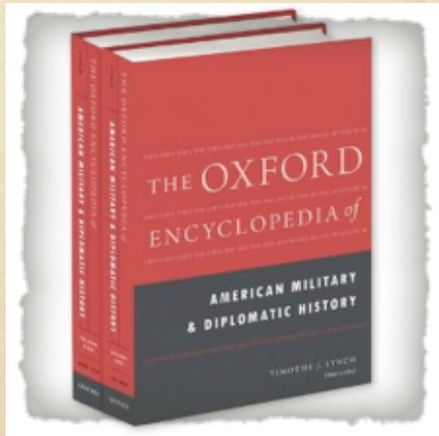
Structured



Specific

General

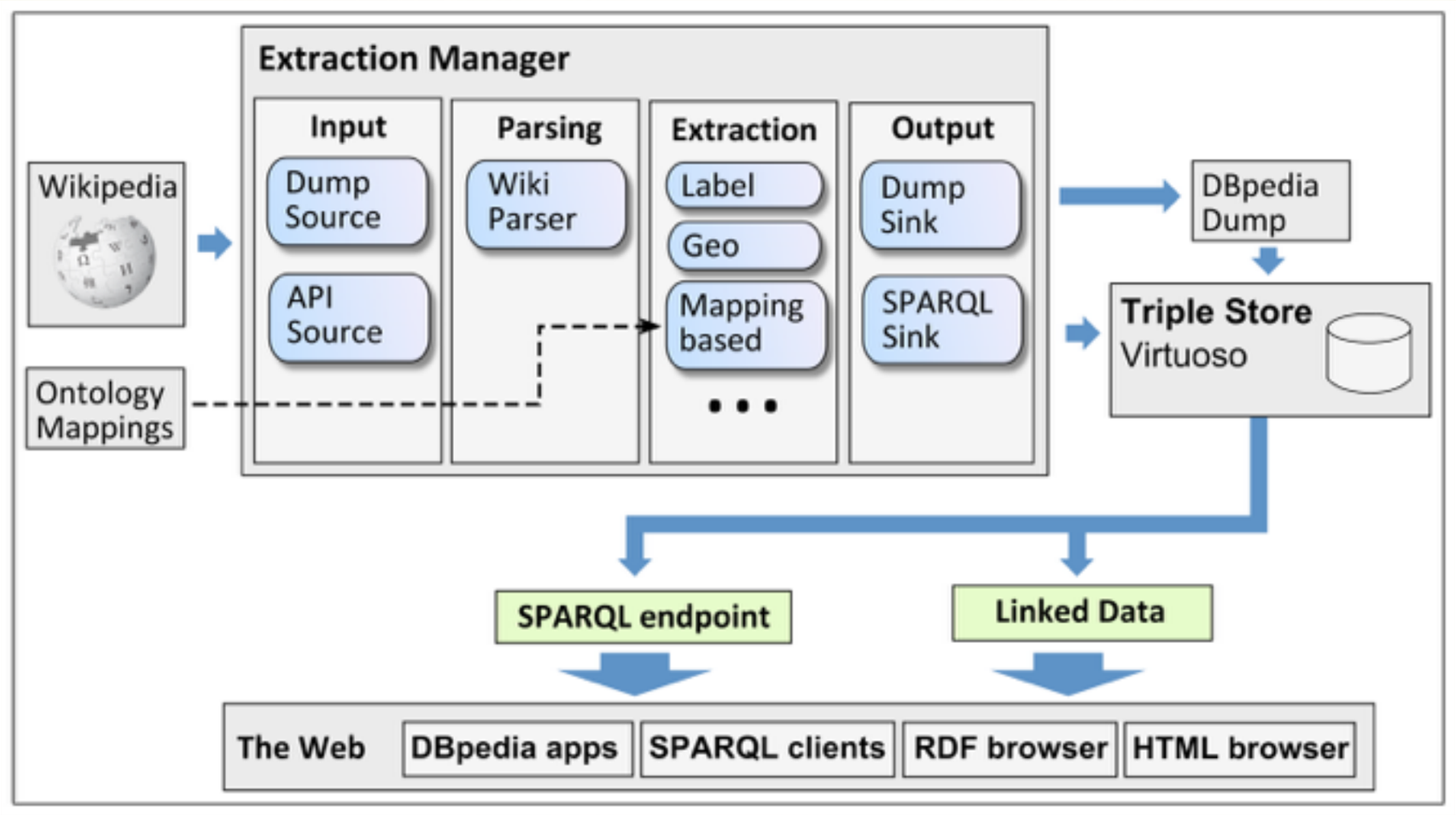




Casual format

DBpedia

- DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia



DBpedia

- One-to-one mapping to wikipedia

http://en.wikipedia.org/wiki/First-order_logic

First-order logic

From Wikipedia, the free encyclopedia

First-order logic is a formal system used in mathematics, philosophy, linguistics, and computer science. It is also known as **first-order predicate calculus**, the **lower predicate calculus**, **quantification theory**, and predicate logic. First-order logic uses quantified variables over (non-logical) objects. This distinguishes it from propositional logic which does not use quantifiers.

A theory about some topic is usually first-order logic together with a specified domain of discourse over which the quantified variables range, finitely many functions which map from that domain into it, finitely many predicates defined on that domain, and a recursive set of axioms which are believed to hold for those things. Sometimes "theory" is understood in a more formal sense, which is just a set of sentences in first-order logic.

The adjective "first-order" distinguishes first-order logic from higher-order logic in which there are predicates having predicates or functions as arguments, or in which one or both of predicate quantifiers or function quantifiers are permitted.[1] In first-order theories, predicates are often associated with sets. In interpreted higher-order theories, predicates may be interpreted as sets of sets.

There are many deductive systems for first-order logic that are sound (all provable statements are true in all models) and complete (all statements which are true in all models are provable). Although the logical consequence relation is only semidecidable, much progress has been made in automated theorem proving in first-order logic. First-order logic also satisfies several metalogical theorems that make it amenable to analysis in proof theory, such as the Löwenheim–Skolem theorem and the compactness theorem.

First-order logic is the standard for the formalization of mathematics into axioms and is studied in the foundations of mathematics. Mathematical theories, such as number theory and set theory, have been formalized into first-order axiom schemas such as Peano arithmetic and Zermelo–Fraenkel set theory (ZF) respectively.

No first-order theory, however, has the strength to describe fully and categorically structures with an infinite domain, such as the natural numbers or the real line. Categorical axiom systems for these structures can be obtained in stronger logics such as second-order logic.

For a history of first-order logic and how it came to dominate formal logic, see José Ferreirós (2001).

Contents [hide]

1 Introduction
2 Syntax
  2.1 Alphabet
    2.1.1 Logical symbols
    2.1.2 Non-logical symbols
  2.2 Formation rules
    2.2.1 Terms
    2.2.2 Formulas
    2.2.3 Notational conventions
  2.3 Free and bound variables
  2.4 Examples
    2.4.1 Ordered abelian groups
    2.4.2 Loving relation
3 Semantics
  3.1 First-order structures
  3.2 Evaluation of truth values
  3.3 Validity, satisfiability, and logical consequence
  3.4 Algebraizations
  3.5 First-order theories, models, and elementary classes
  3.6 Empty domains

http://dbpedia.org/page/First-order_logic

About: First-order logic

An Entity of Type : Logic105664069, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

First-order logic is a formal system used in mathematics, philosophy, linguistics, and computer science. It is also known as first-order predicate calculus, the lower predicate calculus, quantification theory, and predicate logic.

Property	Value

Resource Description Framework

- A general method for conceptual description or modeling of information that is implemented in web resources.
- Make statements about web resources in the form of subject-predicate-object expression.

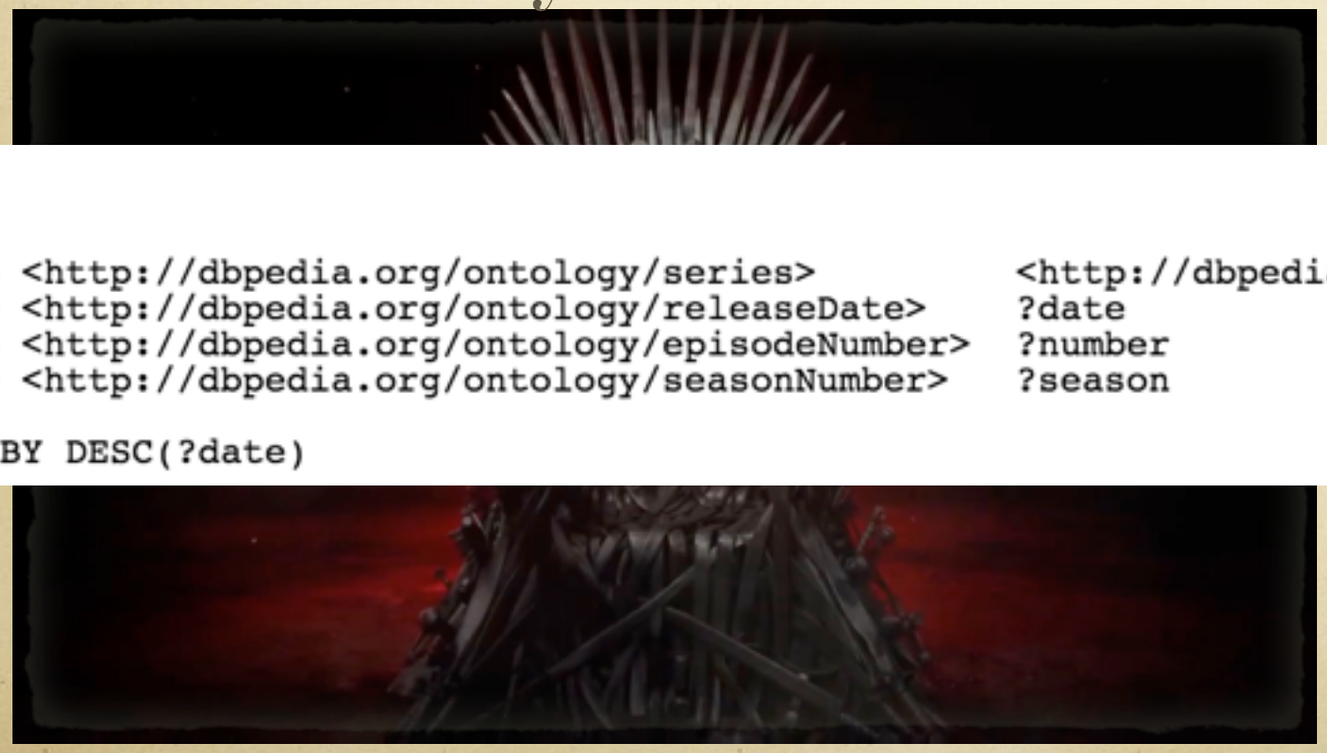
There is a Person identified by http://www.w3.org/People/EM/contact#me, whose name is Eric Miller, whose email address is e.miller123(at)example (changed for security purposes), and whose title is Dr.

- Subject: "http://www.w3.org/People/EM/contact#me"
- The objects are:
  - "Eric Miller" (with a predicate "whose name is"),
  - mailto:e.miller123(at)example (with a predicate "whose email address is"), and
  - "Dr." (with a predicate "whose title is").
- The predicates also have URIs. For example, the URI for each predicate:
  - "whose name is" is http://www.w3.org/2000/10/swap/pim/contact#fullName,
  - "whose email address is" is http://www.w3.org/2000/10/swap/pim/contact#mailbox,
  - "whose title is" is http://www.w3.org/2000/10/swap/pim/contact#personalTitle.

- RDF triples can be expressed:
  - http://www.w3.org/People/EM/contact#me, http://www.w3.org/2000/10/swap/pim/contact#fullName, "Eric Miller"
  - http://www.w3.org/People/EM/contact#me, http://www.w3.org/2000/10/swap/pim/contact#mailbox, mailto:e.miller123(at)example
  - http://www.w3.org/People/EM/contact#me, http://www.w3.org/2000/10/swap/pim/contact#personalTitle, "Dr."
  - http://www.w3.org/People/EM/contact#me, http://www.w3.org/1999/02/22-rdf-syntax-ns#type, http://www.w3.org/2000/10/swap/pim/contact#Person

DBpedia

- Revolutionize Wikipedia Search
  - "Tell me all the episodes of Game of Thrones" rank them by released date.



```
SELECT *
  WHERE
    {
      ?e <http://dbpedia.org/ontology/series>          <http://dbpedia.org/resource/Game_of_Thrones>  .
      ?e <http://dbpedia.org/ontology/releaseDate>     ?date                                          .
      ?e <http://dbpedia.org/ontology/episodeNumber>   ?number                                        .
      ?e <http://dbpedia.org/ontology/seasonNumber>    ?season
    }
  ORDER BY DESC(?date)
```

e	date	number	season
http://dbpedia.org/resource/Oathkeeper	2014-04-27+02:00	4	4
http://dbpedia.org/resource/Breaker_of_Chains	2014-04-20+02:00	3	4
http://dbpedia.org/resource/The_Lion_and_the_Rose	2014-04-13+02:00	2	4
http://dbpedia.org/resource/Two_Swords_(Game_of_Thrones)	2014-04-06+02:00	1	4
http://dbpedia.org/resource/Mhysa	2013-06-09+02:00	10	3
http://dbpedia.org/resource/The_Rains_of_Castamere	2013-06-02+02:00	9	3
http://dbpedia.org/resource/Second_Sons	2013-05-19+02:00	8	3
http://dbpedia.org/resource/The_Bear_and_the_Maiden_Fair	2013-05-12+02:00	7	3
http://dbpedia.org/resource/The_Climb_(Game_of_Thrones)	2013-05-05+02:00	6	3
http://dbpedia.org/resource/Kissed_by_Fire	2013-04-28+02:00	5	3
http://dbpedia.org/resource/And_Now_His_Watch_Is_Ended	2013-04-21+02:00	4	3
http://dbpedia.org/resource/Walk_of_Punishment	2013-04-14+02:00	3	3
http://dbpedia.org/resource/Dark_Wings,_Dark_Words	2013-04-07+02:00	2	3
http://dbpedia.org/resource/Valar_Dohaeris	2013-03-31+02:00	1	3
http://dbpedia.org/resource/Valar_Morghulis	2012-06-03+02:00	10	2
http://dbpedia.org/resource/Blackwater_(Game_of_Thrones)	2012-05-27+02:00	9	2
http://dbpedia.org/resource/The_Prince_of_Winterfell	2012-05-20+02:00	8	2
http://dbpedia.org/resource/A_Man_Without_Honor	2012-05-13+02:00	7	2
http://dbpedia.org/resource/The_Old_Gods_and_the_New	2012-05-06+02:00	6	2
http://dbpedia.org/resource/The_Ghost_of_Harrenhal	2012-04-29+02:00	5	2
http://dbpedia.org/resource/Garden_of_Bones	2012-04-22+02:00	4	2
http://dbpedia.org/resource/What_Is_Dead_May_Never_Die	2012-04-15+02:00	3	2
http://dbpedia.org/resource/The_Night_Lands	2012-04-08+02:00	2	2
http://dbpedia.org/resource/The_North_Remembers	2012-04-01+02:00	1	2
http://dbpedia.org/resource/Fire_and_Blood_(Game_of_Thrones)	2011-06-19+02:00	10	1
http://dbpedia.org/resource/Baelor	2011-06-12+02:00	9	1
http://dbpedia.org/resource/The_Pointy_End	2011-06-05+02:00	8	1
http://dbpedia.org/resource/You_Win_or_You_Die	2011-05-29+02:00	7	1
http://dbpedia.org/resource/A_Golden_Crown	2011-05-22+02:00	6	1
http://dbpedia.org/resource/The_Wolf_and_the_Lion	2011-05-15+02:00	5	1
http://dbpedia.org/resource/Cripples,_Bastards,_and_Broken_Things	2011-05-08+02:00	4	1
http://dbpedia.org/resource/Lord_Snow	2011-05-01+02:00	3	1
http://dbpedia.org/resource/The_Kingsroad	2011-04-24+02:00	2	1
http://dbpedia.org/resource/Winter_Is_Coming	2011-04-17+02:00	1	1

DBpedia

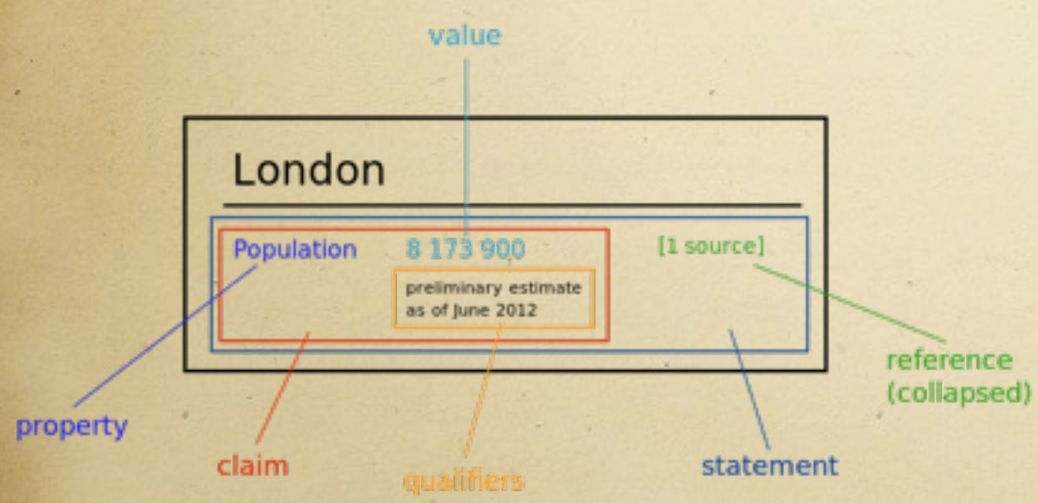
- A lot of other applications
  - http://wiki.dbpedia.org/Applications
- Available in multiple languages
- Downloadable

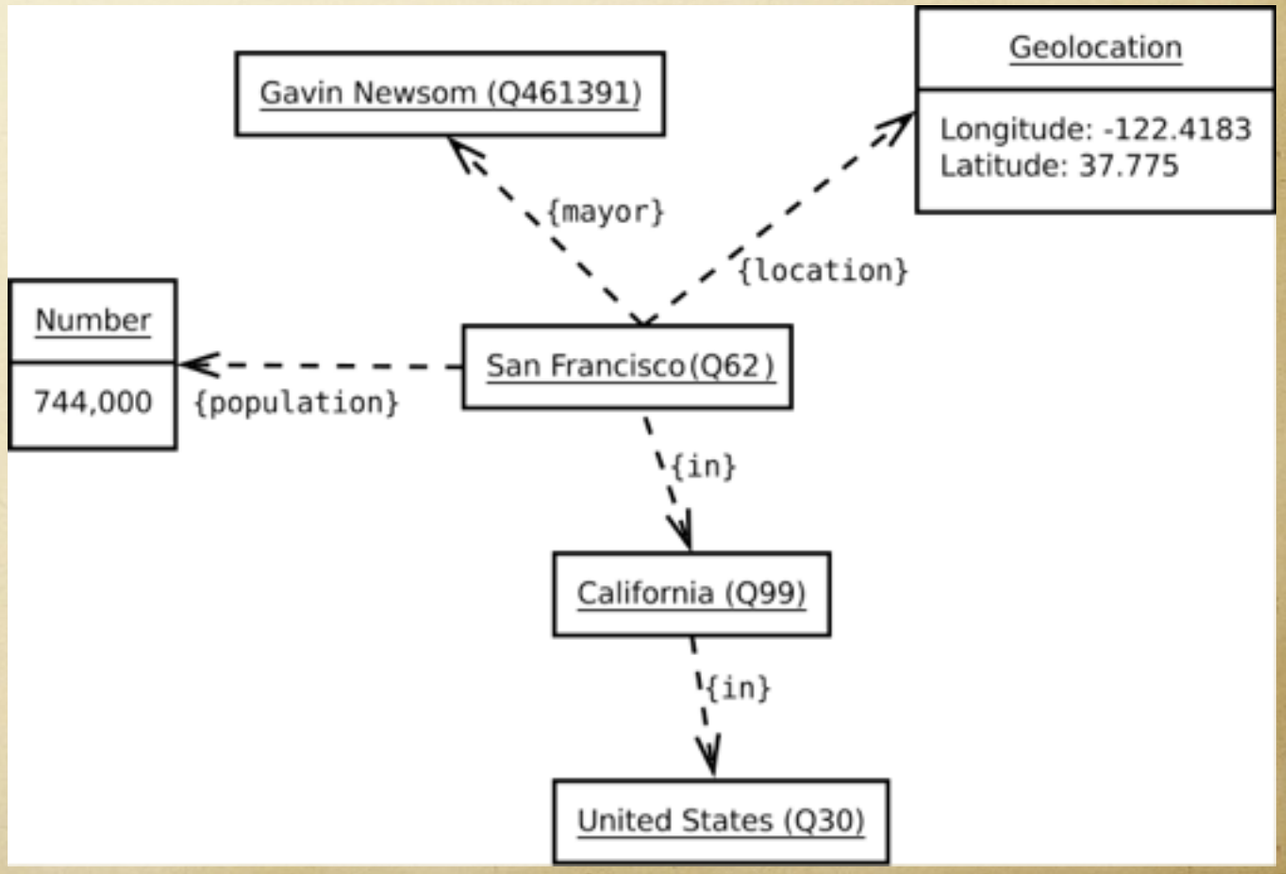
Knowledge Base

- Source of knowledge: internet, human input
- Structure: Graph = Node + Edge
  - RDF: subject-predicate-object
  - Node: entity
  - Edge: relation

WikiData

- Very similar as DBpedia
  - link to more source
  - act as knowledge base for Wikimedia





Wait, wait…

- Knowledge base, structured data organized in graph, DBpedia, Wikidata, Freebase.
- But…

Bring me a cup of hot water

- **Need low level knowledge**

- find a cup
- a cup has *containable* affordance
- cup is empty
- find tape, fill in water
- find microwave
- heat it up

ConceptNet

- A semantic network containing lots of things computers should know about the world.

ConceptNet contains everyday basic knowledge:

learn —*MotivatedByGoal*→ knowledge

You would learn because you want knowledge.

Cultural knowledge:

saxophone —*UsedFor*→ jazz

A saxophone is used for jazz.

a cup has *containable* affordance

And scientific knowledge:

semantic role —*HasContext*→ linguistics

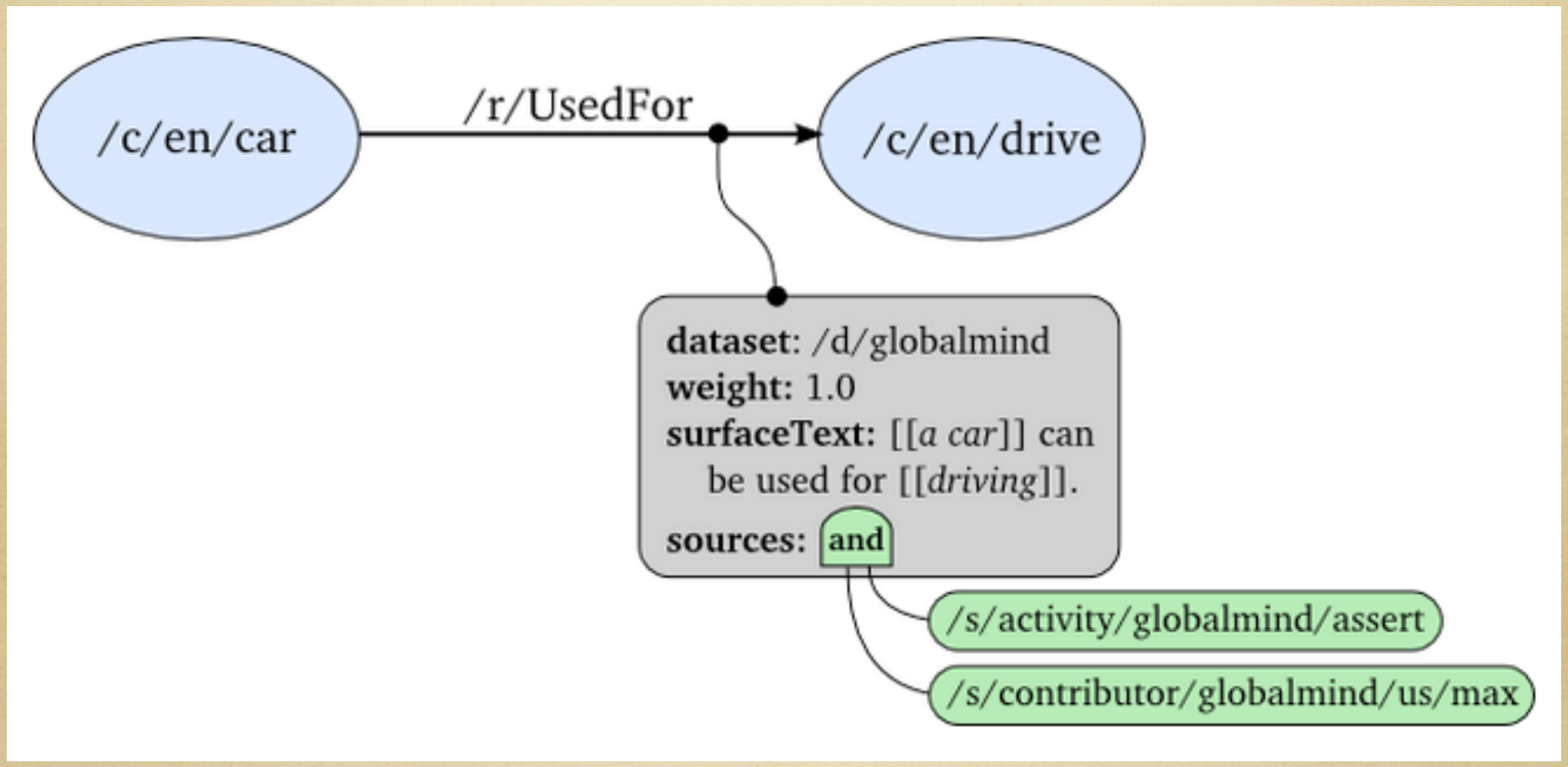
"Semantic role" is a term in linguistics.

It would not adequately represent human knowledge if it didn't contain other languages besides English, as well:

本 —*MadeOf*→ 紙

本は紙でできている。(A book is made of paper.)

ConceptNet


/c/en/car
/r/UsedFor
/c/en/drive
dataset: /d/globalmind
weight: 1.0
surfaceText: [[a car]] can be used for [[driving]].
sources: and
/s/activity/globalmind/assert
/s/contributor/globalmind/us/max

ConceptNet

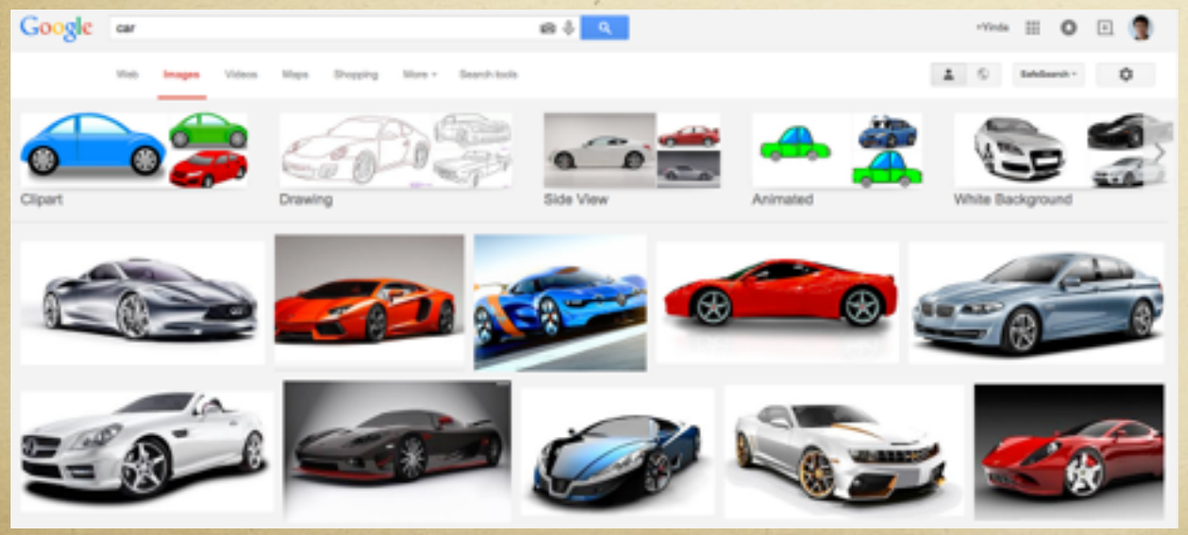
- Free to download
- Provide API to:
  - Retrieve the data for particular nodes and edges
  - Query for edges with given properties
  - Measure and query the semantic distance between nodes

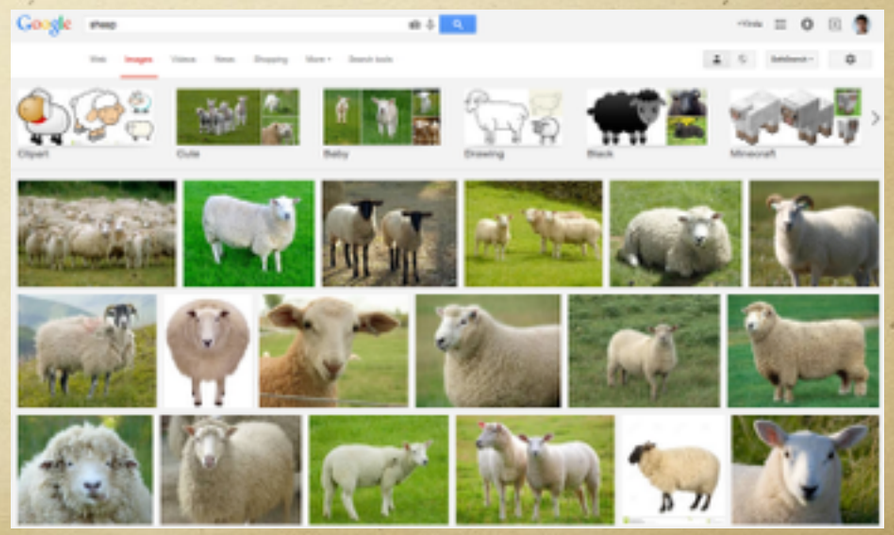
So far…

- There are lexical knowledge base for both high-level and low-level knowledge ready online.
- To connect the knowledge with computer vision, we need visual knowledge base.
  - Not as explicit as language
    - "A car can be used for driving"

Never Ending Image Learner

- Learn from image searching engine (the weak association between image and text)
  - what a car looks like?
  - know that sheep are white





Never Ending Image Learner

- NEIL is a computer program
  - Run 24h per day, 7 days per week
  - Automatically extract visual knowledge from internet data
  - Learn to see

  - Learn common sense

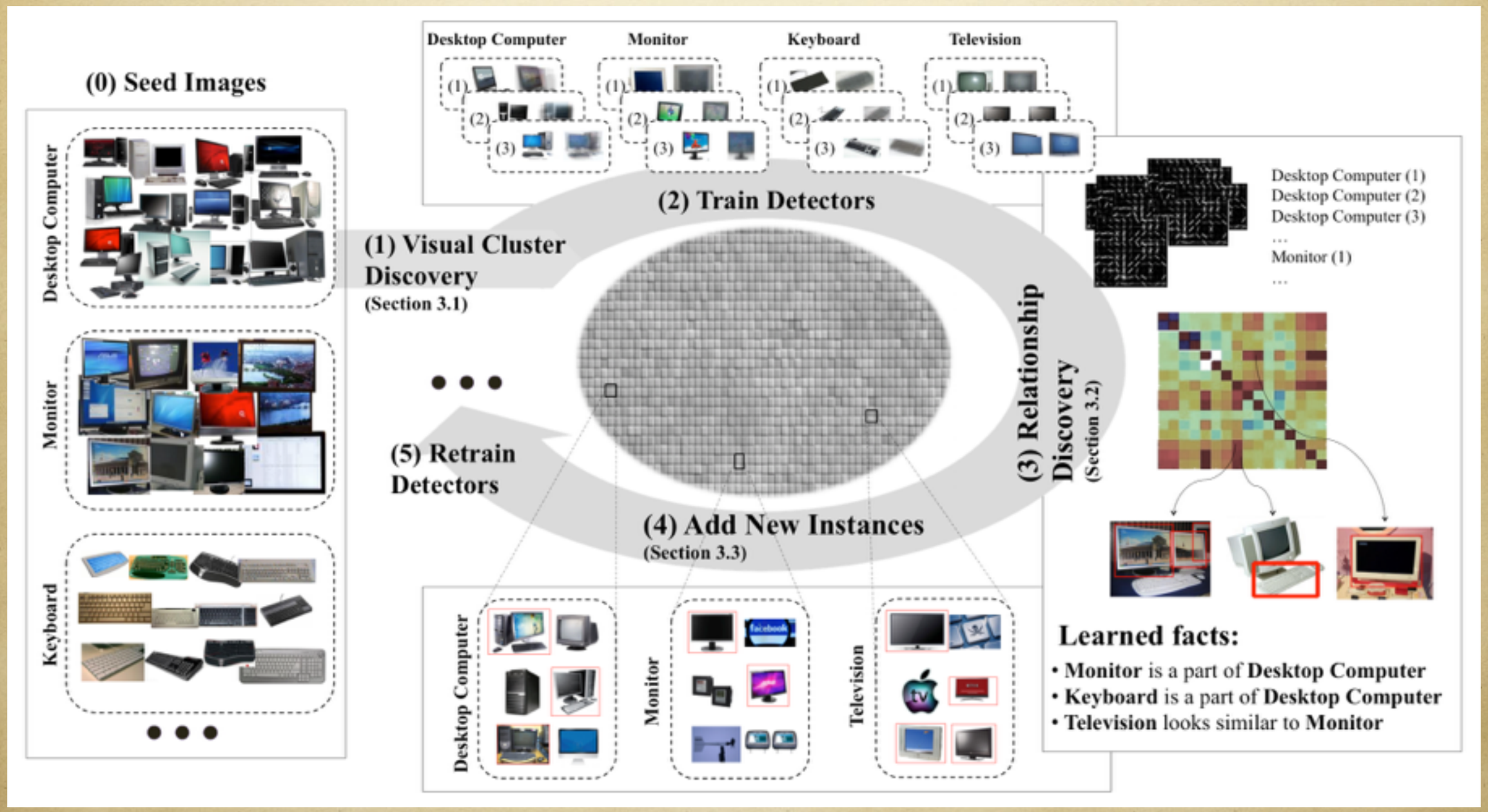
Airbus_330 can be a kind of / look similar to Airplane.

Car can have a part Wheel.

Bamboo_forest can be / can have Vertical_lines.

Trading_floor can be / can have Crowded.

Never Ending Image Learner



Never Ending Image Learner

- Seeding Classifier via Google Image Search
  - scene, attribute classifier; object, attribute detector.
  - Directly train scene and attribute classifier on downloaded images.
  - However, fail for object and attribute detector
    - Outlier, Polysemy, Visual diversity, Localization

Never Ending Image Learner

- Seeding Classifier via Google Image Search
  - Train exemplar-LDA for each image
  - Run detection on all images
  - Get top K windows with high scores from multiple detectors
  - Clustering with ELDA score vector
  - Train classifier for each cluster

Never Ending Image Learner

- Seeding Classifier via Google Image Search

(a) Google Image Search for "tricycle"

(b) Sub-category Discovery

Never Ending Image Learner

- Extract Relationships
  - Object-Object Relationships:
    - Partonomy: Eye is a part of Baby.
    - Taxonomy:BMW 320 is a kind of Car.
    - Similarity: Swan looks similar to Goose.

Never Ending Image Learner

- Extract Relationships
  - Build co-occurrence matrix
  - Get co-occurred object pairs
  - Learn relationship in terms of mean and variance of relative positive, aspect ratio, score, size.

Never Ending Image Learner

- Object-Attribute Relationships
  - “Pizza has Round Shape”, “Sunflower is Yellow”
- Scene-Object Relationships
  - “Bus is found in Bus depot”
- Scene-Attribute Relationships
  - “Ocean is Blue”

Never Ending Image Learner

- Discover new instance and retrain

$$\phi_i(x) + \sum_{i,j \in \mathcal{R}_O \bigcup \mathcal{R}_A} \phi_j(x_l)\psi_{i,j}(x, x_l) + \sum_{i,k \in \mathcal{R}_S} \omega_k(x)$$

object detector

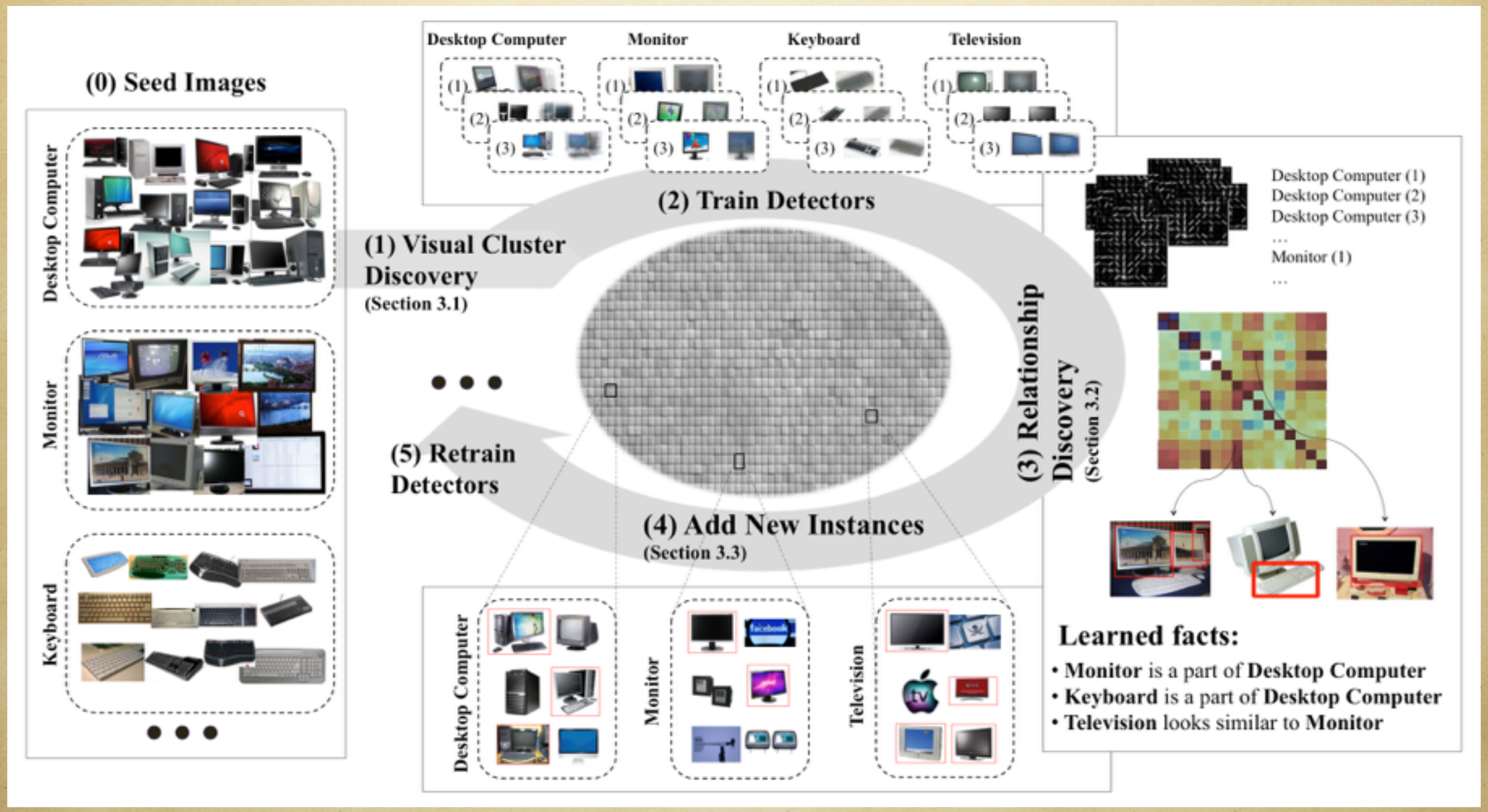
binary relationship

all related objects and attributes

scene classifier

all related scenes

Never Ending Image Learner



Never Ending Image Learner

- Bootstrapping
  - Words: NELL (never ending language learning)
  - Images: ImageNet, SUN, Google Image Search

Table 1. mAP performance for scene classification on 12 categories.

	mAP
Seed Classifier (15 Google Images)	0.52
Bootstrapping (without relationships)	0.54
NEIL Scene Classifiers	0.57
NEIL (Classifiers + Relationships)	**0.62**

Table 2. mAP performance for object detection on 15 categories.

	mAP
Latent SVM (50 Google Images)	0.34
Latent SVM (450 Google Images)	0.28
Latent SVM (450, Aspect Ratio Clustering)	0.30
Latent SVM (450, HOG-based Clustering)	0.33
Seed Detector (NEIL Clustering)	0.44
Bootstrapping (without relationships)	0.45
NEIL Detector	0.49
NEIL Detector + Relationships	**0.51**

Hey, it's about time…

- to fix the annoying problem

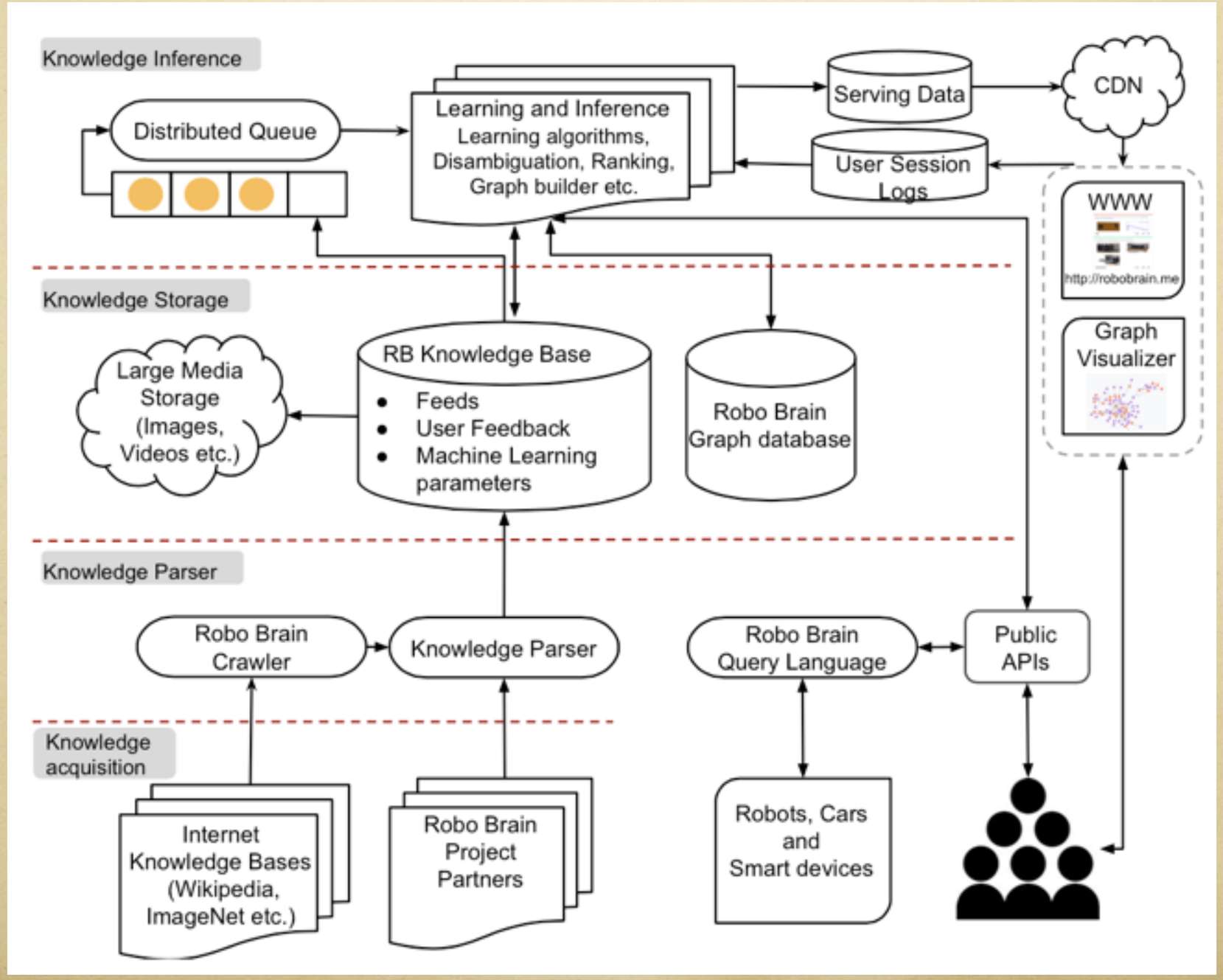
  Bring me a cup of hot water

- Design a robot with knowledge base

RoboBrain

- A large-scale knowledge engine for robot
- Build a knowledge base similar as ConceptNet
- More diverse edges
- Edges have beliefs
  - measure the confidence of learned relations
  - labelled by crowd-sourced feedback

RoboBrain


Knowledge Inference
Distributed Queue
Learning and Inference
Learning algorithms, Disambiguation, Ranking, Graph builder etc.
Serving Data
CDN
User Session Logs
WWW
http://robobrain.me
Graph Visualizer
Knowledge Storage
Large Media Storage (Images, Videos etc.)
RB Knowledge Base
Feeds
User Feedback
Machine Learning parameters
Robo Brain Graph database
Knowledge Parser
Robo Brain Crawler
Knowledge Parser
Robo Brain Query Language
Public APIs
Knowledge acquisition
Internet Knowledge Bases (Wikipedia, ImageNet etc.)
Robo Brain Project Partners
Robots, Cars and Smart devices

RoboBrain

- How to build knowledge base?
  - again, graph represented in triplets

Word	An english word represented as an ASCII string.
DeepFeature	Feature function trained with a Deep Neural Network
Image	2D RGB Image.
PointCloud	3D point cloud

IsTypeOf	human *IsTypeOf* a mammal.
HasAppearance	floor *HasAppearance* as follows (this image).
CanPerformAction	human *CanPerformAction* cutting.
SpatiallyDistributedAs	location of human is *SpatiallyDistributedAs*.

(StandingHuman, Shoe, *CanUse*)
(Grasping, DeepFeature23, *UsesFeature*)
(StandingHuman, $\mathcal{N}(\mu, \Sigma)$, *SpatiallyDistributedAs*)

RoboBrain

- Knowledge acquisition

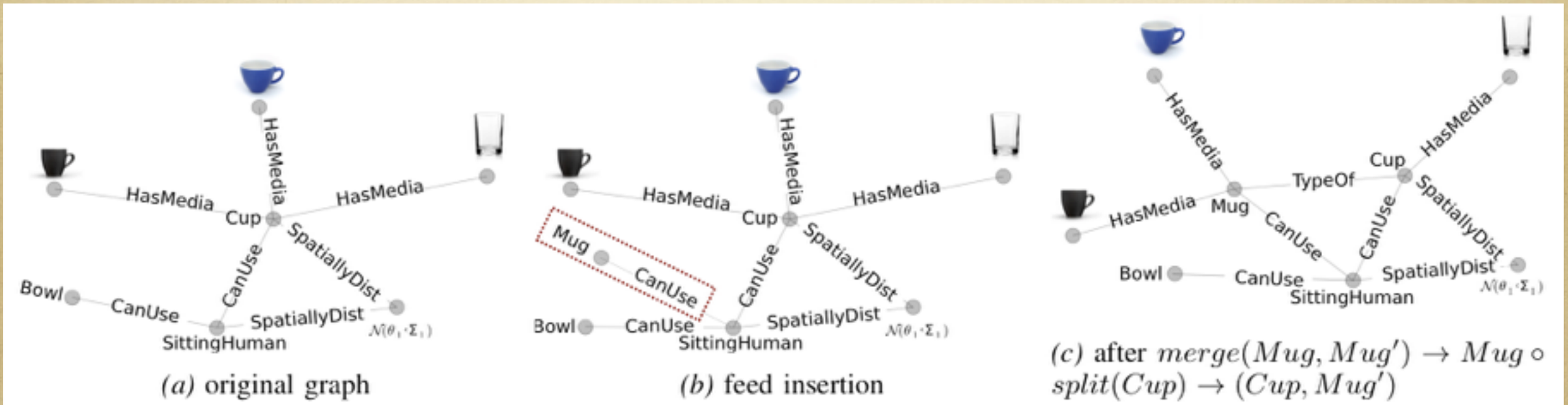
$$G = (V, E) \quad + \quad (v_1^1, v_2^1, \ell^1) \ldots (v_1^N, v_2^N, \ell^N) \quad \rightarrow \quad G' = (V', E')$$

Original Database | New Feeds | New Database

$$V' = v_1^1 \cup v_2^1 \cup \ldots \cup v_1^N \cup v_2^N \cup V$$
$$E' = (v_1^1, v_2^1, \ell^1) \cup \ldots \cup (v_1^N, v_2^N, \ell^N) \cup E$$

RoboBrain

- Merge and Split



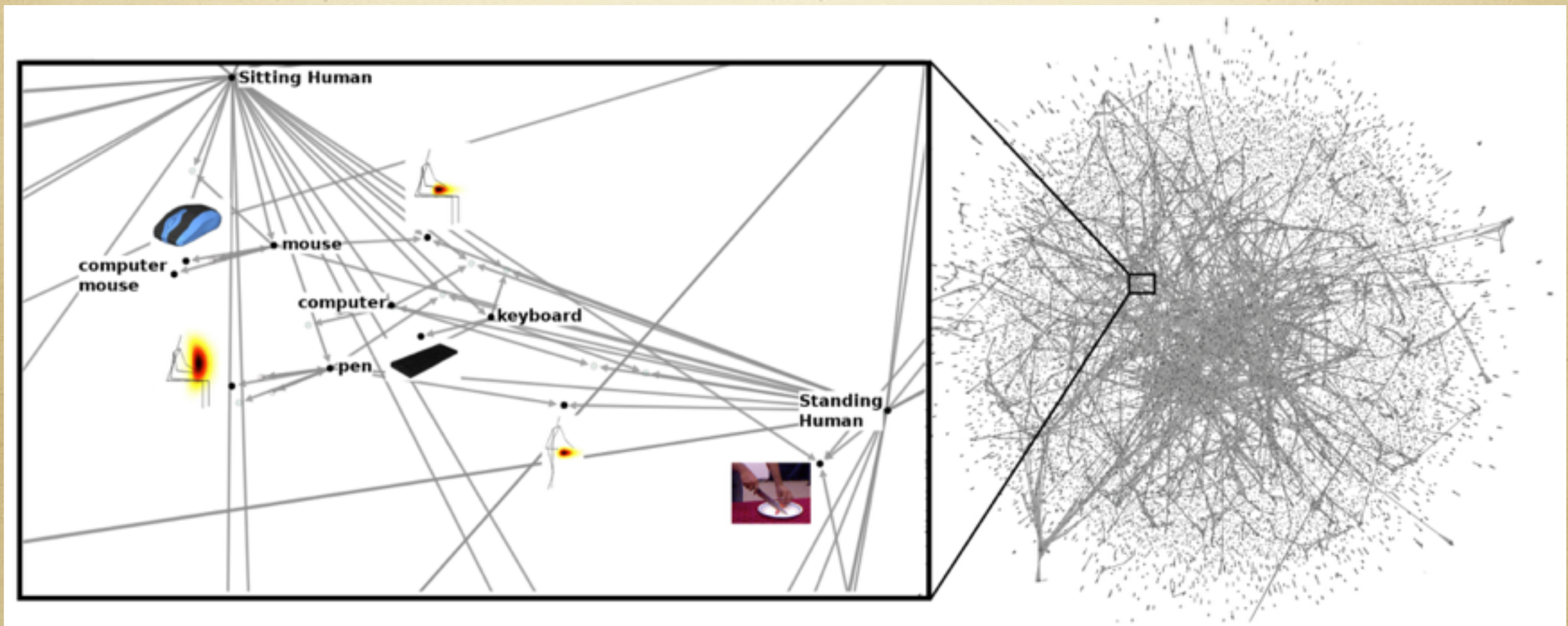
(a) original graph

(b) feed insertion

(c) after $merge(Mug, Mug') \rightarrow Mug \circ split(Cup) \rightarrow (Cup, Mug')$

RoboBrain

- Visualization of Knowledge Base



50K nodes, 100K edges

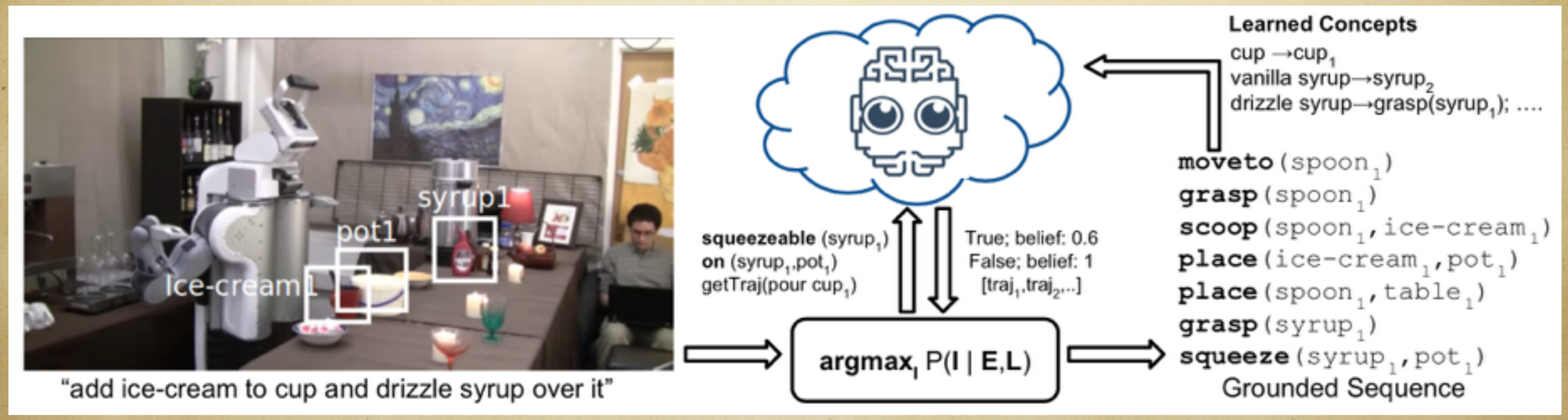
RoboBrain

- Grounding a natural language sentence
  - “fill a cup with water”

moveto(cup_{01}); *grasp*(cup_{01}); *moveto*($sink_{01}$); *keep*(cup_{01}, *on*, $sink_{01}$); *toggle*($sink\_knob_{01}$); *wait*(); *toggle*($sink\_knob_{01}$);

RoboBrain

- Grounding a natural language sentence
  - appearance, affordance, possible action, associated trajectory, manipulation feature



RoboBrain

- Support action planning

$squeeze(syrup_1, pot_1)$

- *grasping pr2 syrup*$_1$: Robot is grasping the syrup.
- *squeezeable syrup*$_1$: Syrup bottle should be squeezeable.
- *on syrup*$_1$ *pot*$_1$: Syrup bottle should be placed directly above the pot.

$$\textit{squeezeable syrup}_1 \;=\; \texttt{len fetch}\ (\texttt{u}\{\texttt{name} : \text{`}\texttt{syrup}'\}) \rightarrow [\text{`}\texttt{HasAffordance}'] \rightarrow (\texttt{v}\{\texttt{name} : \text{`}\texttt{squeezeable}'\}) \;>\; 0$$

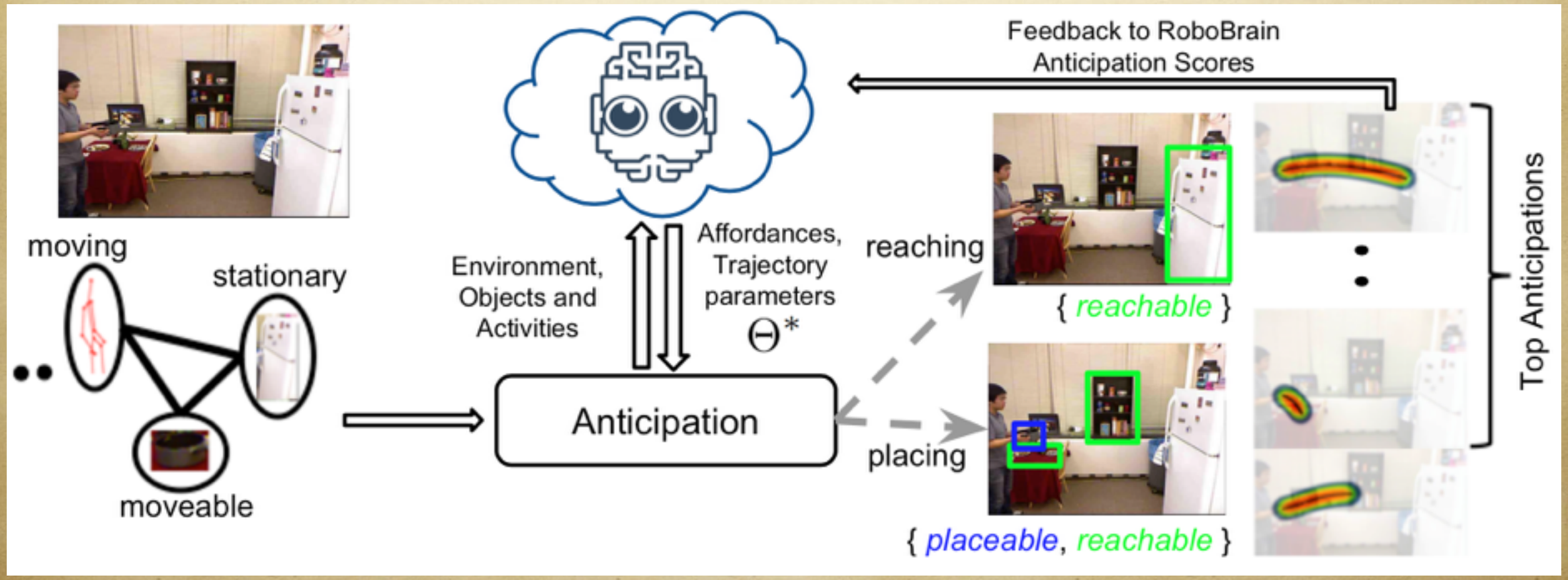
RoboBrain

- Transfer action primitives to trajectory

$$\begin{aligned} \mathit{pour\ cup}_{01} \ &= \ \texttt{fetch}\ (\{\texttt{name} : \text{`}\texttt{pour}'\}) \rightarrow \\ &(\texttt{v}\{\texttt{name} : \text{`}\texttt{HasTrajectory}'\}) \leftarrow (\{\texttt{name} : \text{`}\texttt{cup}'\}) \end{aligned}$$

RoboBrain

- Other application
  - anticipating human activity



RoboBrain

- Summary
  - a knowledge base integrates knowledge about physical world that robots live in.
  - share knowledge to support complicated tasks
    - natural language grounding
    - activity prediction

Can we do more?

- So far, we know how to reuse learned knowledge.
- Can we generalize the learned knowledge to understand what we never seen before?

edible

Zero-shot Affordance Prediction

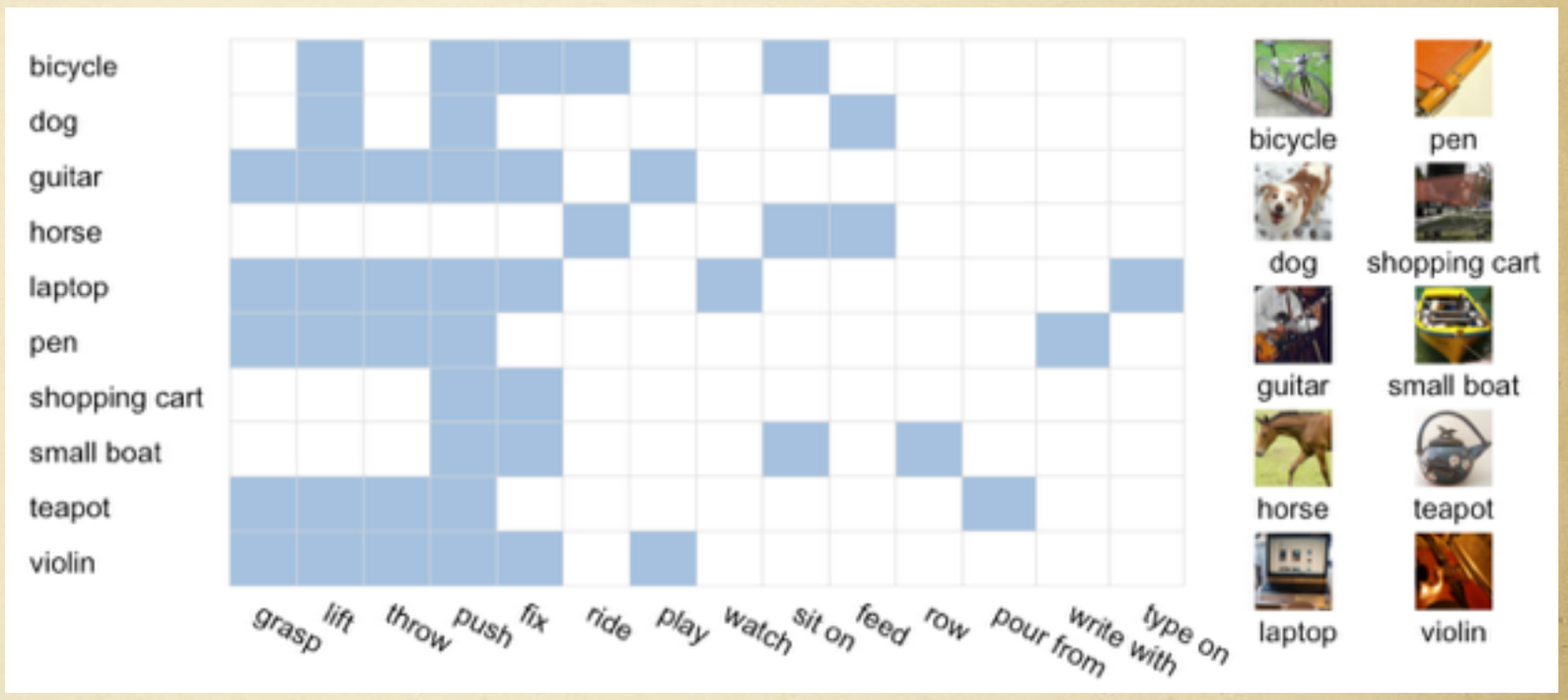
- Idea
  - affordance, attribute, human interaction are highly correlated

Zero-shot Affordance Prediction

- Learning the knowledge base: choose 40 objects (Stanford 40 Action Database)
  - Nodes (Entities):
  - Attribute:
    - visual: 33 per-trained classifiers, “round”, “shiny”
    - physical: weight, size, from FreeBase, Amazon
    - categorical: 22 from WordNet, “animal”, “vehicle”

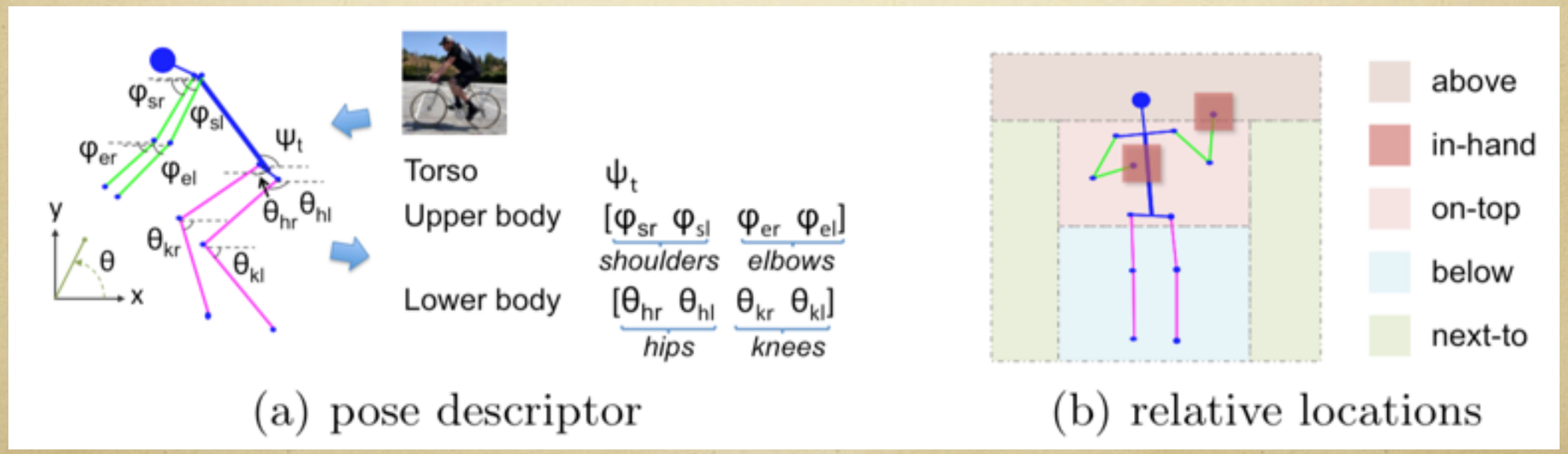
Zero-shot Affordance Prediction

- Nodes
  - Attributes
  - Affordance
    - choose 14 from Stanford 40 Action
    - manual labeling for 40 objects
    - on average, 4.25 per object



Zero-shot Affordance Prediction

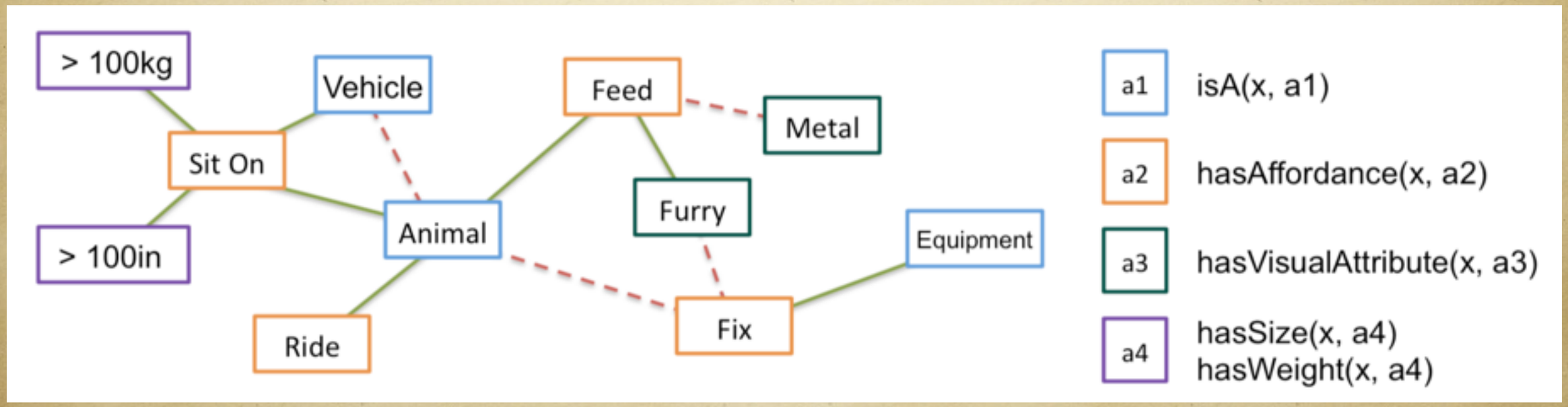
- Nodes:
  - Human pose: cluster centroids of descriptor.
  - Human object relative position



(a) pose descriptor (b) relative locations

Zero-shot Affordance Prediction

- Learn a Markov Logic Network (MRF) to represent the relationships between nodes
- Use training data to build such relationships



Zero-shot Affordance Prediction

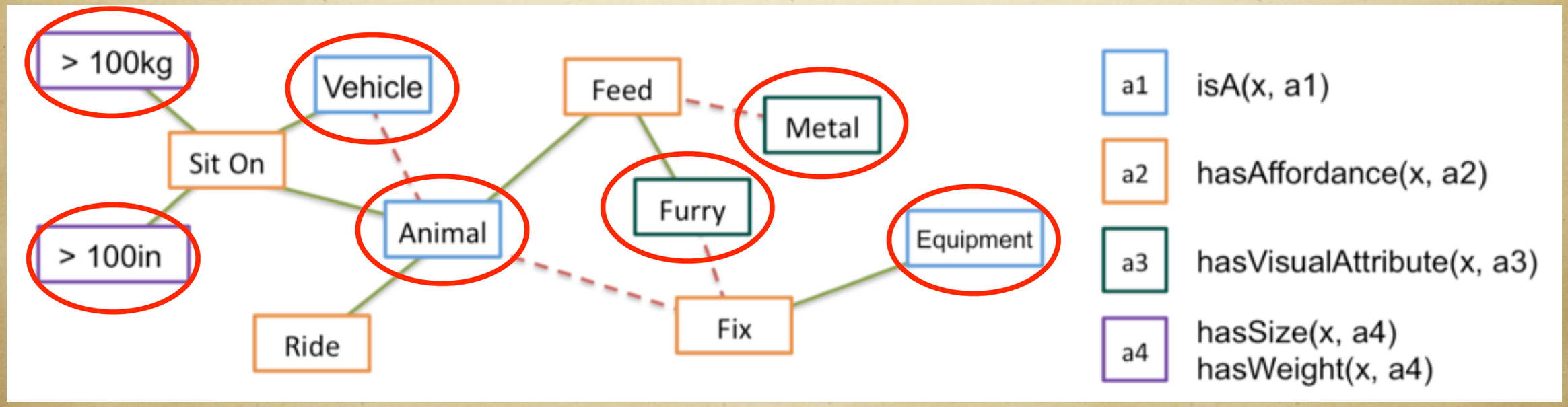
- Zero-shot prediction:
  - choose 22 objects that are semantically similar as the 40 training objects.
  - sample 50 images per objects as testing set.

Zero-shot Affordance Prediction

- Zero-shot prediction:
  - Estimating visual attributes: run classifiers
  - Inferring:
    - Categorical attributes: learn regression from image feature and VA
    - Physical attributes: regression from image feature

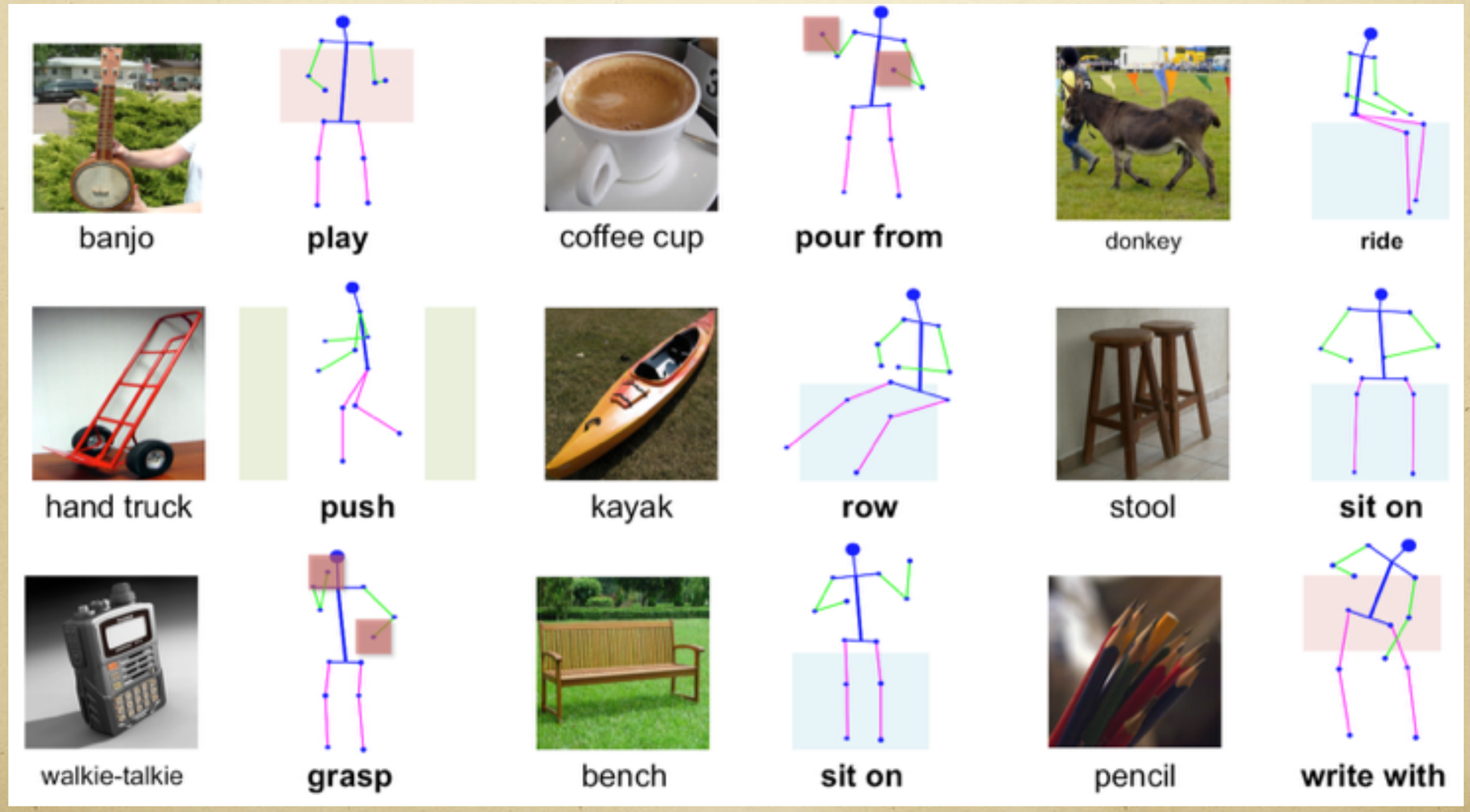
Zero-shot Affordance Prediction

- Zero-shot prediction:
  - Now, we have confidence on attribute nodes. Run belief propagation on MRF, we get confidence on affordance nodes.



Zero-shot Affordance Prediction

- Zero-shot prediction:



Zero-shot Affordance Prediction

- Zero-shot prediction:

Table 1. Performance of Zero-shot Affordance Prediction (measured in mAUC)

Method	**L1-LR** [9]	χ^2**-SVM** [25]	**Ours**
base features (BF)	0.7858	-	-
visual attributes (VA)	0.7525	0.7533	0.7432
categorical & physical (CP)	0.7919	0.7924	0.8234
combined (VA+CP)	0.8006	0.7985	**0.8409**

Table 2. Performance of Estimating Human Poses (in Hamming distance)

Method	nearest neighbor	attributes	affordances	attributes+affordances
Distance	0.928	1.027	0.630	**0.527**

Zero-shot Affordance Prediction

- Prediction from human pose:

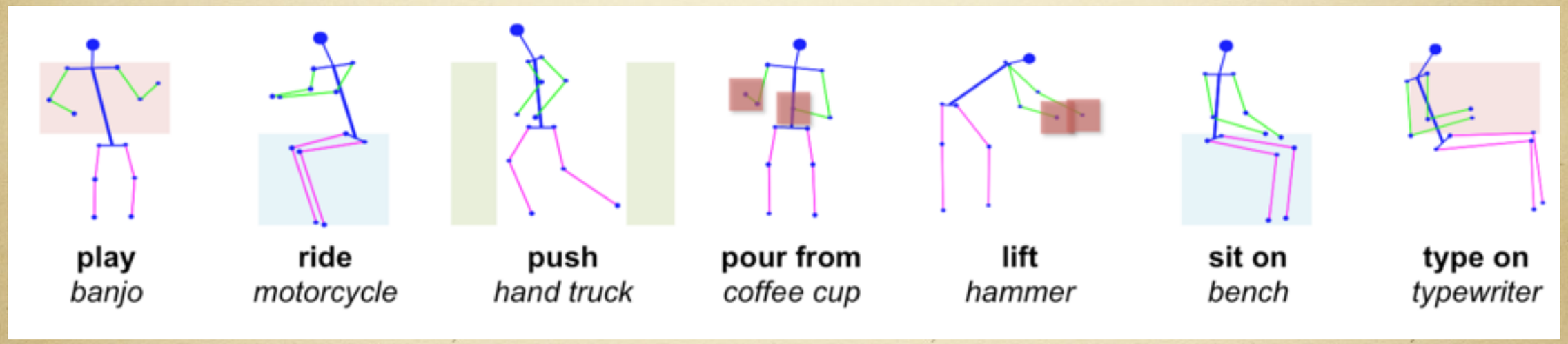


Table 3. Predicting Actions and Objects from Human-Object Interactions

Method	Action	Object
human poses	50.4%	46.2%
poses + locations	81.2%	64.5%

Zero-shot Affordance Prediction

- Robust to partial observation:

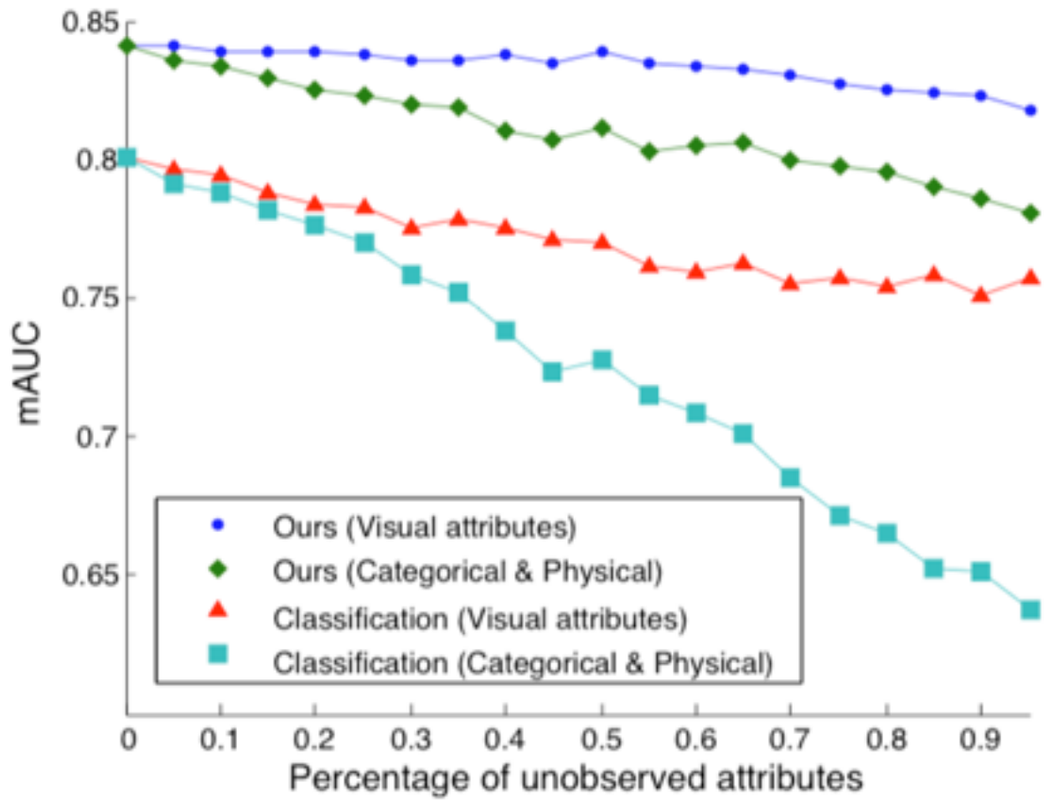


Fig. 12: **Performance variations against partial observation.** The x-axis denotes the percentage of unobserved evidence. The y-axis denotes the performance (mAUC). The top two curves correspond to our method. The bottom two are the classification-based method. In comparison, the knowledge base representation is more robust against partial observation.

Zero-shot Affordance Prediction

- Question Answering:

Question	Evidence	Query	Top Answers
What do animals look like?	isA(N1, **Animal**)	hasVisualAttribute(N1, x)	hasVisualAttribute(N1, **Leather**) hasVisualAttribute(N1, **Head**) hasVisualAttribute(N1, **Tail**) hasVisualAttribute(N1, **Furry**)
I saw something shiny and metallic. What is it?	hasVisualAttribute(N1, **Shiny**) hasVisualAttribute(N1, **Metal**)	isA(N1, x)	isA(N1, **Instrumentality**) isA(N1, **Device**) isA(N1, **Container**) isA(N1, **Computer**)
Here is a vehicle and it's quite heavy. What can I do with it?	isA(N1, **Vehicle**) hasWeight(N1, W4) **(> 100 kg)**	hasAffordance(N1, x)	hasAffordance(N1, **Ride**) hasAffordance(N1, **Row**) hasAffordance(N1, **SitOn**) hasAffordance(N1, **Fix**)
Tell me how heavy and large a wooden musical instrument is.	isA(N1, **Musical_instrument**) hasVisualAttribute(N1, **Wood**)	hasWeight(N1, x) hasSize(N1, x)	hasSize(N1, D2) **(10-100 in)** hasWeight(N1, W2) **(1-10 kg)**

Summary

- Online knowledge base
  - high-level: DBpedia, Wikidata
  - low-level: ConceptNet
- How to learn visual knowledge base: NEIL
- How to create KB for robot to do complicated tasks: RoboBrain
- How to generalize KB: zero-shot affordance prediction