Announcements

Assignments:

- HW6
  - Due Tomorrow 3/3, 10 pm
- P3
  - Due Thu 3/5, 10 pm
  - Waiving up to two slip days. Submitting by Sat 3/7 doesn’t cost you anything except your break, but you cannot submit after 3/7 at 10pm.

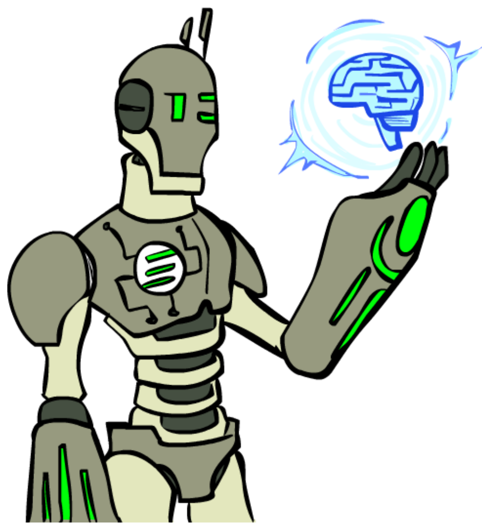
Announcements

Assignments:

- HW6
  - Due Tomorrow 3/3, 10 pm
- P3
  - Due Thu 3/5, 10 pm
  - Waiving up to two slip days. Submitting by Sat 3/7 doesn't cost you anything except your break, but you cannot submit after 3/7 at 10pm.
- HW7
  - Out Tomorrow 3/3. Due Thurs 3/19 at 10pm
- HW8
  - Out Tomorrow 3/17. Due Tue 3/24 at 10pm

AI: Representation and Problem Solving

Knowledge Representation

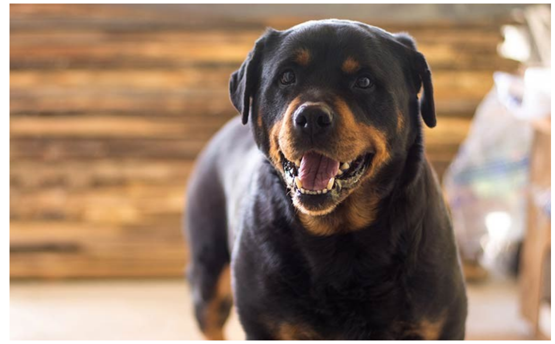


Instructors: Pat Virtue & Stephanie Rosenthal

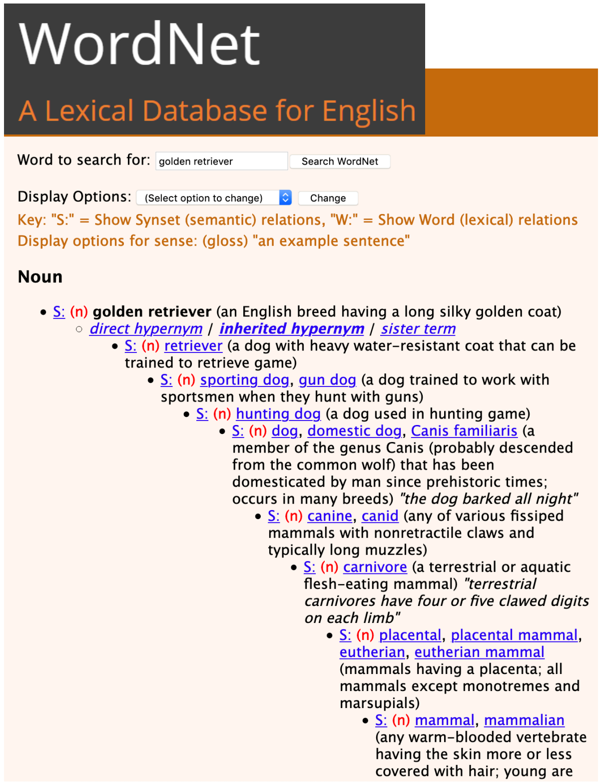
Slide credits: CMU AI, especially Tom Mitchell (NELL)

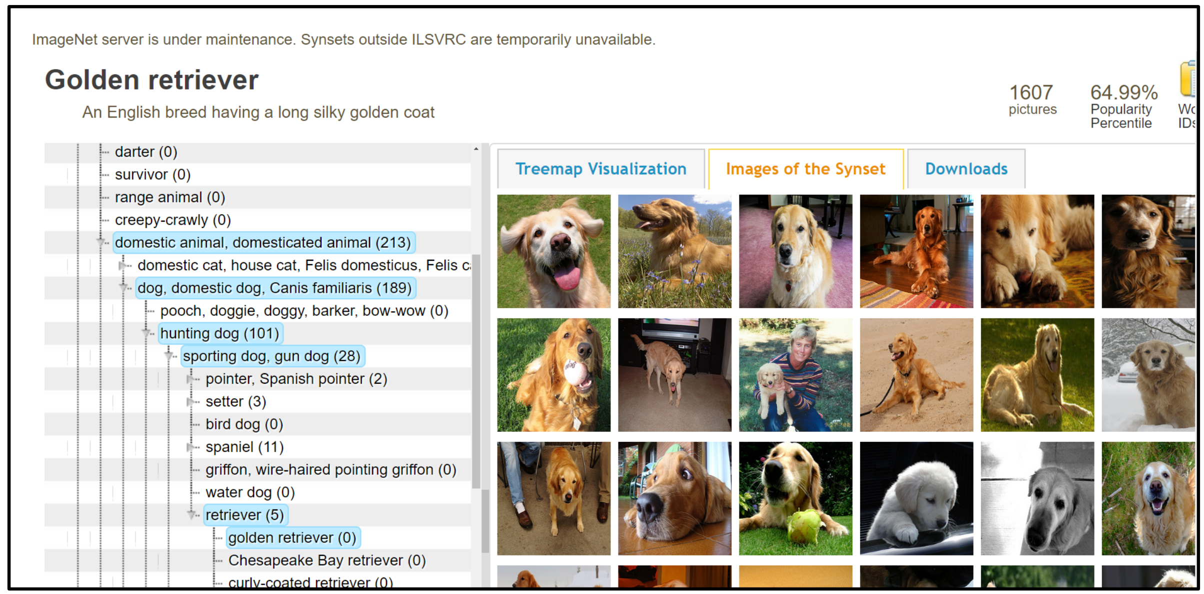
What is this?

What is this?



Ontologies

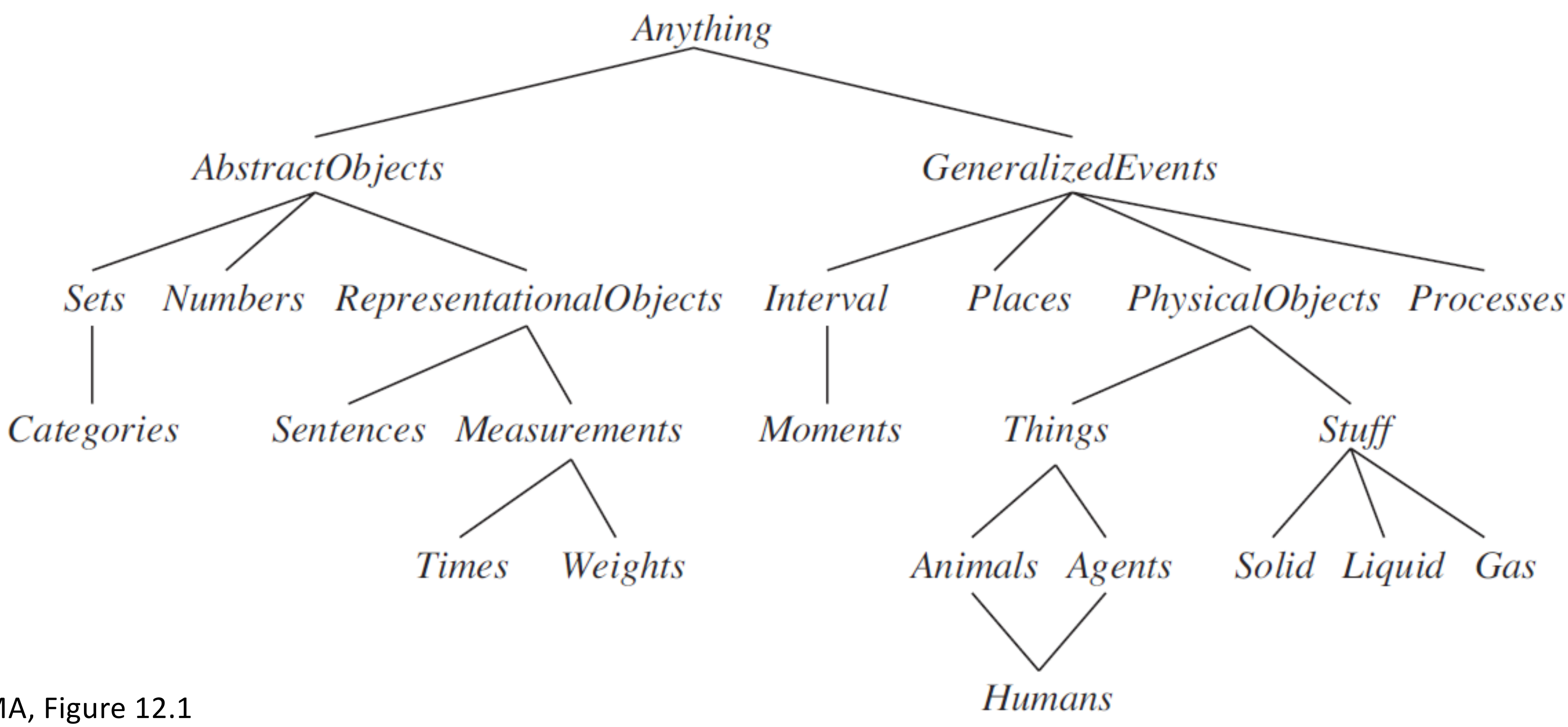




https://wordnet.princeton.edu/

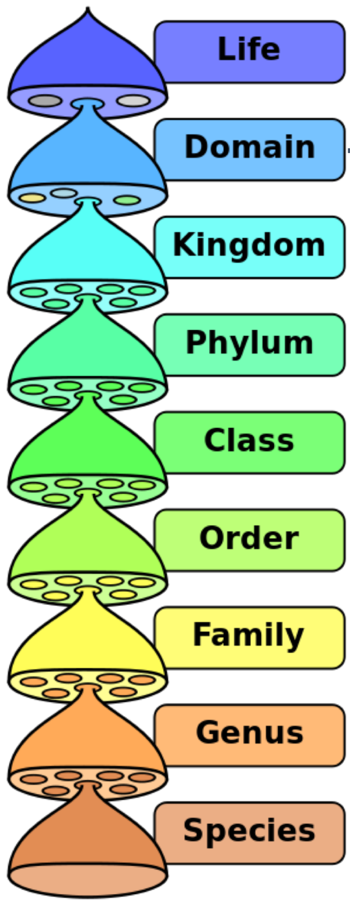
http://www.image-net.org/

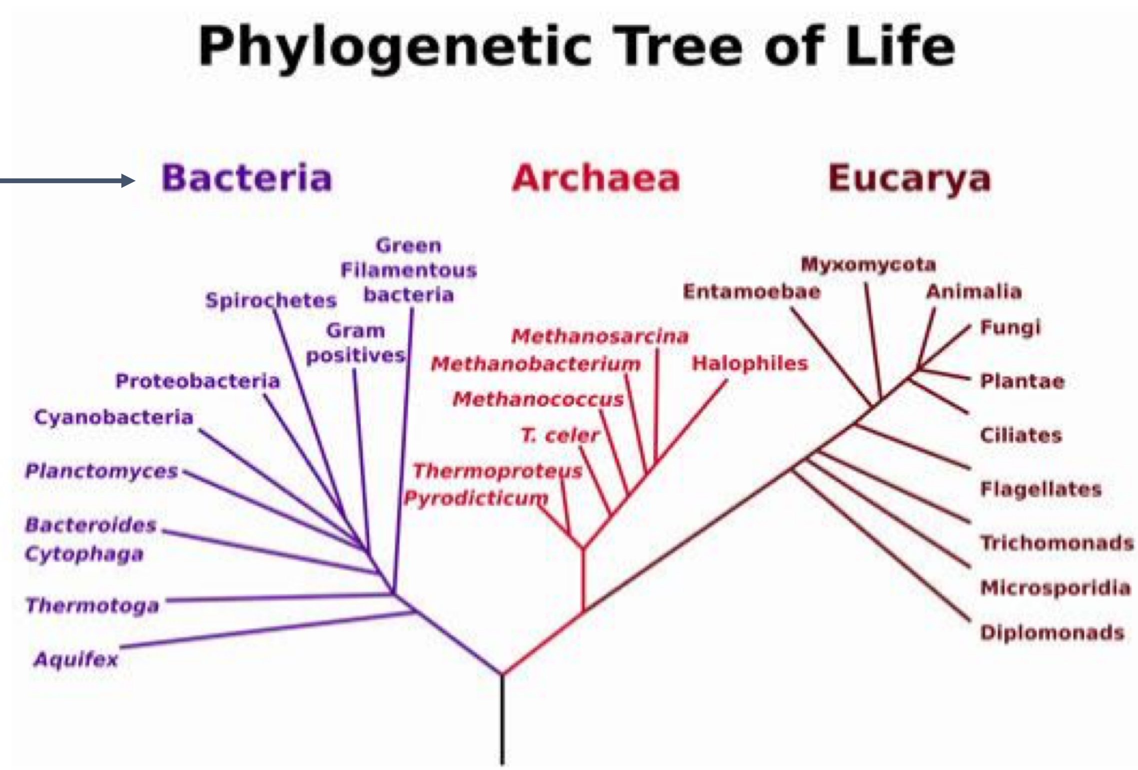
An “upper ontology” of the world



AIMA, Figure 12.1

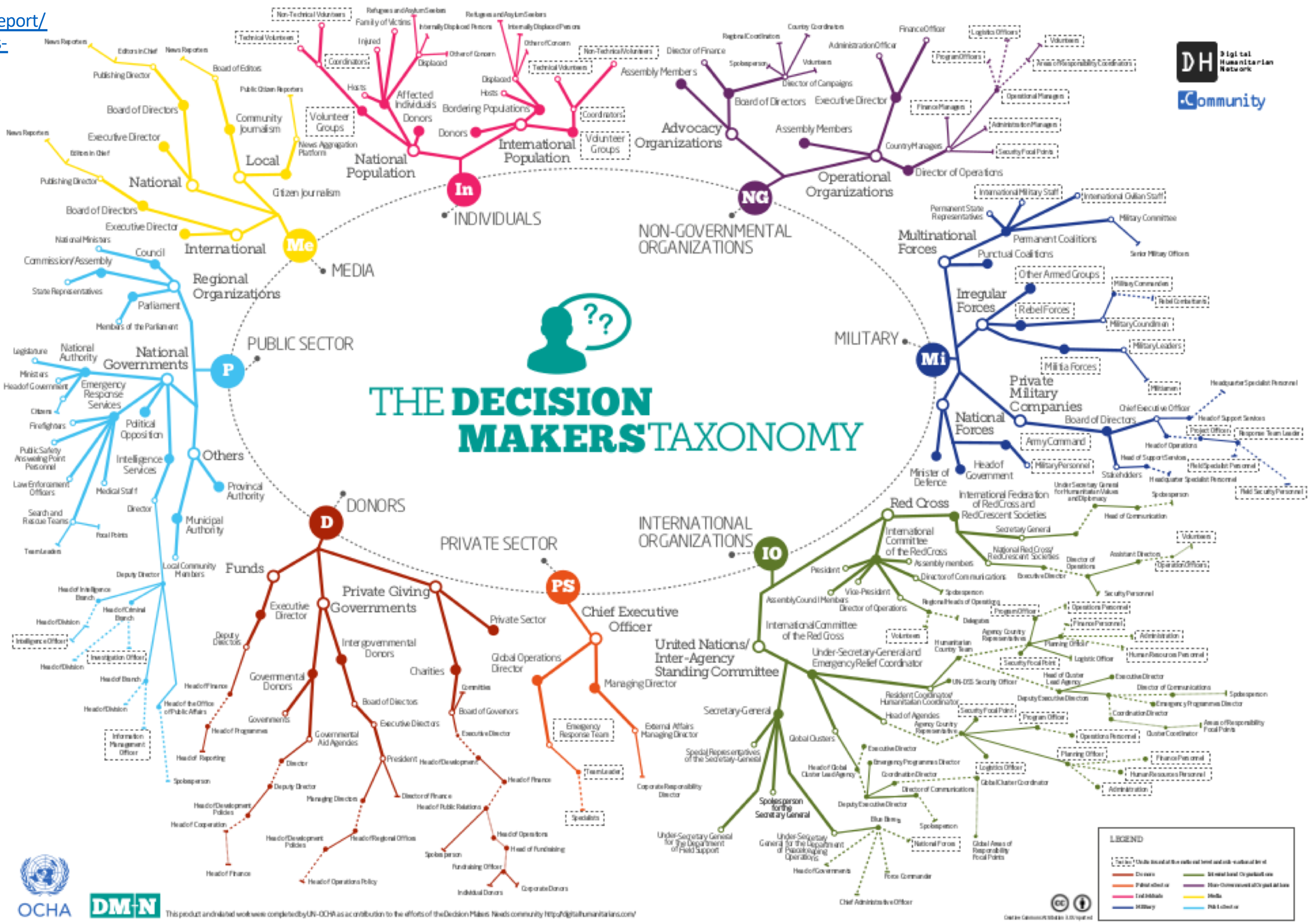
Taxonomic Hierarchies





10 million living and extinct species.

https://reliefweb.int/report/world/decision-makers-taxonomy



VIDEO GAME MOOD TAXONOMY

TAXONOMY CREATION & APPLICATION FROM THE PERSPECTIVE OF USERS

STEPHANIE ROSSI | MLIS

INTRODUCTION

The mood taxonomy, developed by the Game Metadata Research Group, in collaboration with the SIMM, is part of the Video Game Metadata Schema. This project expands upon current research, furthering our understanding of how people percieve and describe the mood of video games and interactive media.

METHOD

Metadata: video game title list with 1500+ entries.

User Study and Exercise: 26 gamers completed questionnaires, were interviewed and asked to apply mood terms to popular video game titles.

RESULTS

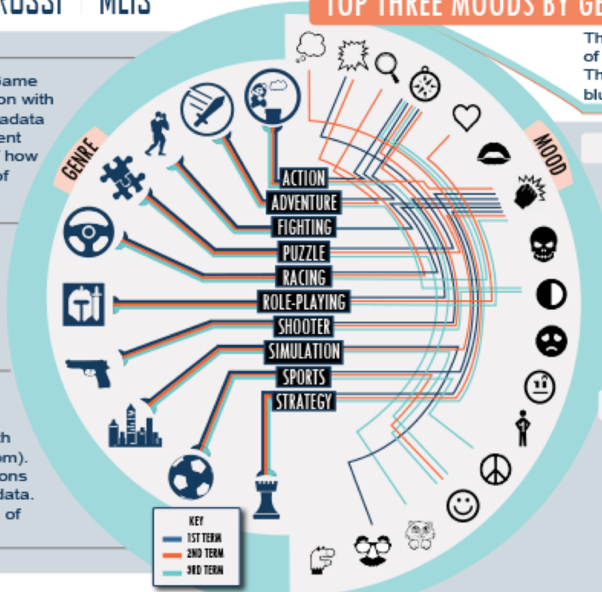
Study Data: interview transcripts and exercise data.

Data Analysis: sample of 300 games with sourced genre and year (from allgame.com).

Term Suggestions: mood term suggestions collected from gamerDNA and interview data.

Mood Clustering: hierarchical clustering of current preferred terms.

TOP THREE MOODS BY GENRE



The diagram to the left explores the genre breakdown of mood terms applied to a sample of 300 games. The top three mood terms are represented by dark blue, orange, and light blue lines, repectively.

CURRENT MOOD TERMS

adventurous, humorous, quirky, aggressive, imaginative, romantic, cute, intense, sad, dark, light-hearted, sarcastic, horror, mysterious, sensual, peaceful, solitary

MOOD TERM SUGGESTIONS

Preferred Terms	Equivalent Terms
• casual*	• charming
• competitive*	• exciting
• epic*	• happy
• friendly/social	• weird
• immersive	• wonder

*Mentioned in both gamerDNA and interview data

MOOD STUDY DATA ANALYSIS

As part of the exercise, participants were asked to apply mood terms to 35 video game titles. Below are the four most recognized video game titles with their applied mood terms.

While a large number of participants applied the same top mood term to *Angry Birds* (light-hearted), *Assassin's Creed III* (adventurous) and *Grand Theft Auto IV* (aggressive) opinion was more split over *Super Mario Bros.*

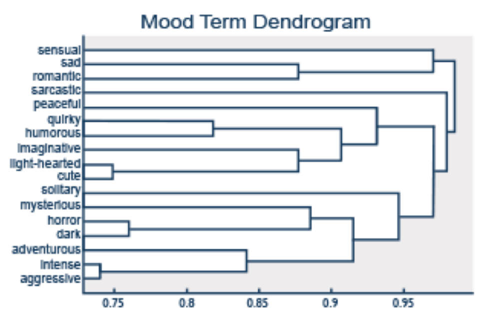
ANGRY BIRDS

ASSASSIN'S CREED III

GRAND THEFT AUTO IV

SUPER MARIO BROS.

MOOD CLUSTER DATA



This graph represents mood terms that appear frequently together in the collected interview data. The results indicate that the mood terms intense and aggressive, light-hearted and cute, and horror and dark are applied often together when describing video game mood. Clustering mood terms will allow us to use groups of particular mood terms instead of individual terms to obtain more consistent results.

Acknowledgements: Jin Ha Lee, Carl Gellert, Alison Lane, The Seattle Interactive Media Museum, The Game Metadata Research Group, INFO498/INFX598 Participants, Jaki Parsons, and The Noun Project.

Categories and Objects

First-order logic for ontological representations

Category: Basketball

- Predicate: $Basketball(b)$
- Object for category: $Basketballs$
  - $Member(b, Basketballs)$
    - Notation shortcut: $b \in Basketballs$
  - $Subset(Basketballs, Balls)$
    - Notation shortcut: $Basketball \subset Balls$
- Specific object
  - $Basketball_{12} \in Basketballs$

Reification: converting category predicate into an object

Categories and Objects

Decompositions and Partitions

$Disjoint(\{Animals, Vegetables\})$

$ExhaustiveDecomposition(\ \{Canadians, Americans, Mexicans\},\ \ NorthAmericans\ )$

$Partition(\ \{Canada, United\ States, Mexico\},\quad NorthAmericanCountries\}\ )$

(disjoint and exhaustive decomposition)

Categories and Objects

Parts

$PartOf(Bucharest, Romania)$

$PartOf(Romania, EasternEurope)$

$PartOf(EasterEurope, Europe)$

Transitive

$PartOf(x, y) \wedge PartOf(y, z) \Rightarrow PartOf(x, z)$

Reflexive

$PartOf(x, x)$

Categories and Objects

Measurements

Number are objects

Units are typically functions to convert number constants to measurements

$$Length(L_1) = Inches(1.5) = Centimeters(3.81)$$

Piazza Poll 1

Which of these measurement statements makes sense?

Select ALL that apply.

A) $Diameter(Basketball)$
B) $Diameter(Basketball_{12})$
C) $Weight(Apple)$
D) $Weight(Apple_1 \wedge Apple_2 \wedge Apple_3)$
E) None of the above

Categories and Objects

Bunches of Things and Stuff

$BunchOf(\{Apple_1, Apple_2, Apple_3\})$

Things

- Countable
- “The” apple, “an” apple

Stuff

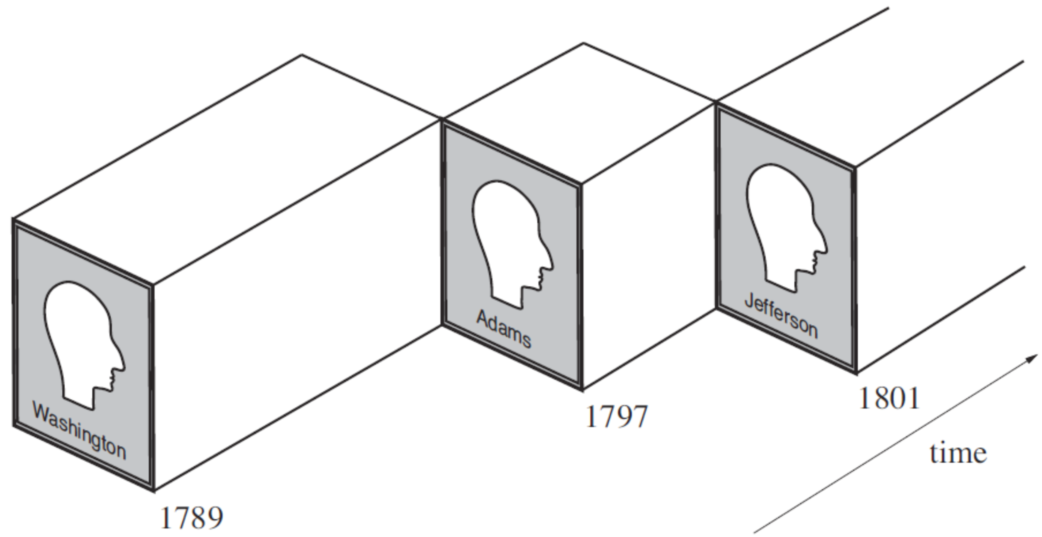
- More of a mass
- “Some” water
- $b \in Butter \wedge PartOf(p, b) \Rightarrow p \in Butter$

Events

How to handle fluents?

$President(USA)$

$President(USA, t)$

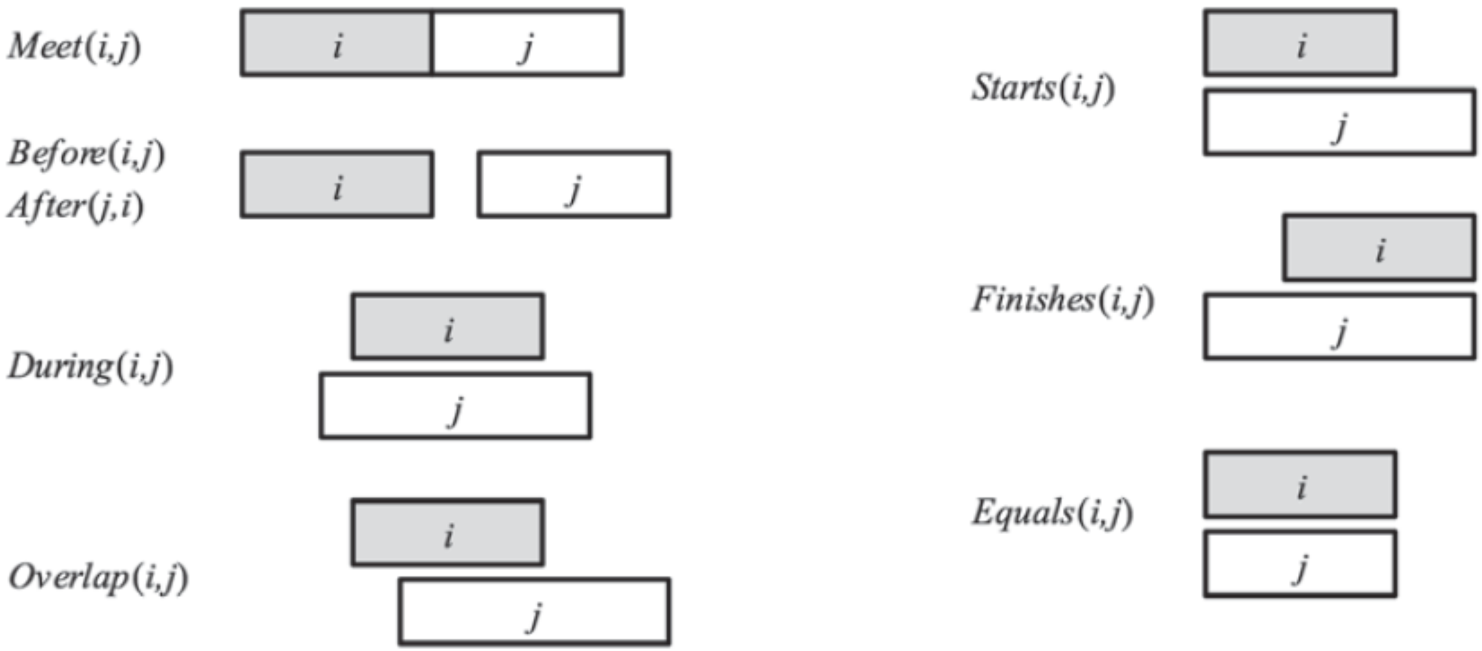


$$During\left(AD1790,\ TimeRange\left(GeorgeWashington, President(USA)\right)\right)$$

AIMA, Figure 12.3

Events

How to handle time?



AIMA, Figure 12.2

Events

How to handle time?

- Meet(i, j): End(i) = Begin(j)
- Before(i, j): End(i) < Begin(j)
- After(j, i): Before(i, j)
- During(i, j): Begin(j) < Begin(i) < End(i) < End(j)
- Overlap(i, j): Begin(i) < Begin(j) < End(i) < End(j)
- Starts(i, j): Begin(i) = Begin(j)
- Finishes(i, j): End(i) = End(j)
- Equals(i, j): Begin(i) = Begin(j) and End(i) = End(j)

Semantic Networks

A graphical representation for some types of knowledge

- Once viewed as an “alternative” to logic (it’s not really)
- The IS-A relation often forms the backbone of a semantic network

Vertebrate

IS-A

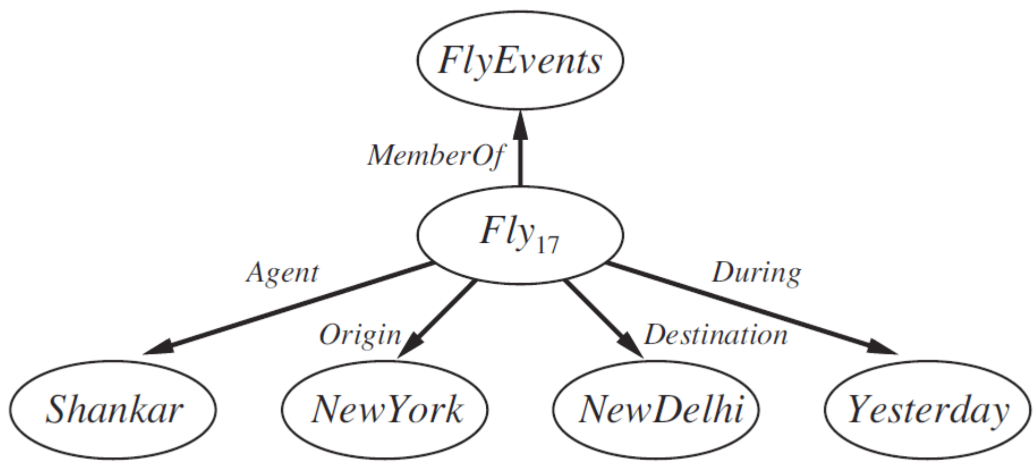
Mammal

IS-A

Elephant

InstanceOf

Clyde



Semantic Networks

Reasoning with default information

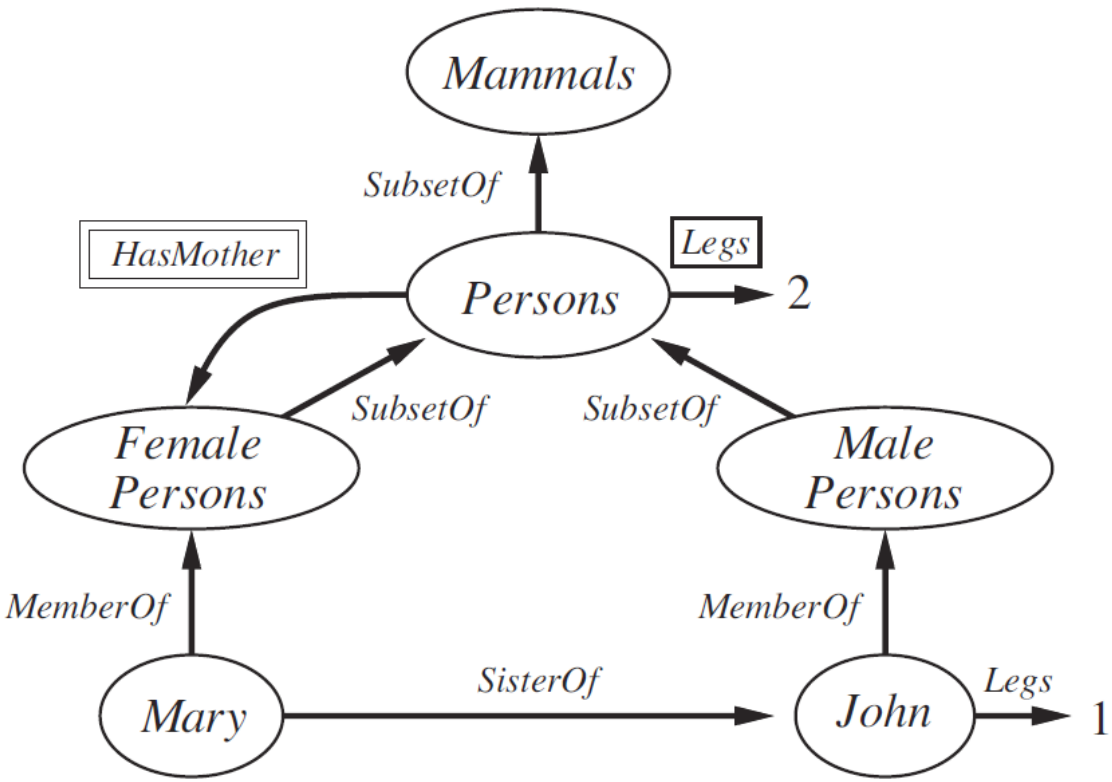
Dog

- Barks
- Has Fur
- Has four legs

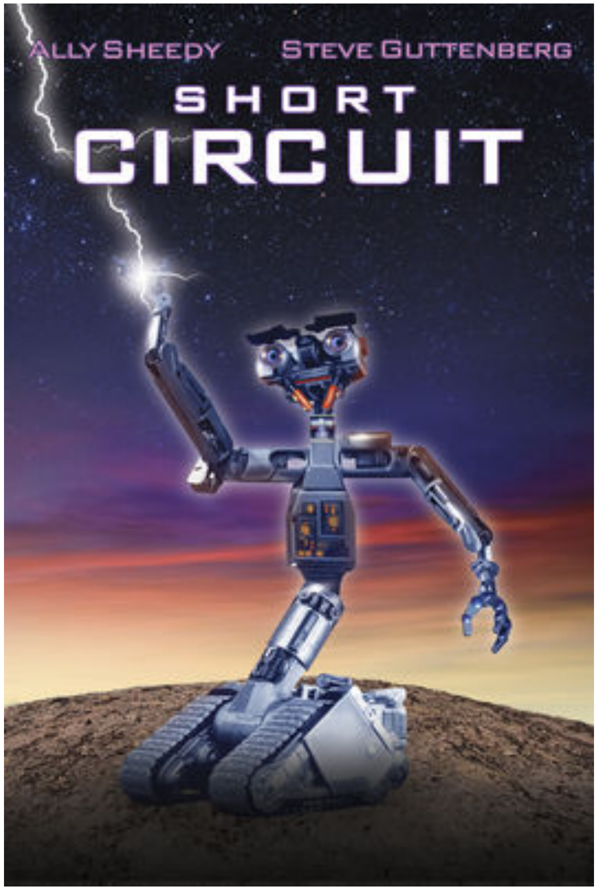
Buster

Semantic Networks

Reasoning with default information



Input, More Input!



https://www.youtube.com/watch?v=x3rhtSToXto

Knowledge Representation in the Wild

- WordNet
- ImageNet
- Wikimedia: Wikipedia, WikiData
- Google Knowledge Graph
- Schema.org
- “Learning a Health Knowledge Graph from Electronic Medical Records”
- The “Semantic Web”
- NELL: Never Ending Language Learning

Knowledge panels in Google search results

The panels are generated from what's called the **Google Knowledge Graph**.

Data comes from Wikipedia, CIA World Factbook, and other online sources.

As of Oct. 2016, held 70 billion facts for 1 billion entities

Google Knowledge Graph API Access

```
import json
import urllib
api_key = open('.api_key').read()
query = 'Taylor Swift'
service_url = 'https://kgsearch.googleapis.com/v1/entities:search'
params = {
    'query': query,
    'limit': 10,
    'indent': True,
    'key': api_key,
}

url = service_url + '?' + urllib.urlencode(params)
response = json.loads(urllib.urlopen(url).read())
for element in response['itemListElement']:
  print element['result']['name'] + ' (' + str(element['resultScore']) + ')'
```

Partial result


```
{ "@type": "EntitySearchResult",
  "result": {
    "@id": "kg:/m/0dl567",
    "name": "Taylor Swift",
    "@type": [
      "Thing",
      "Person"
    ],
    "description": "Singer-songwriter",
    "image": {
      "contentUrl": "https://t1.gstatic.com/images?q=tbn:ANd9GcQm...",
      "url": "https://en.wikipedia.org/wiki/Taylor_Swift",
      "license": "http://creativecommons.org/licenses/by-sa/2.0"
    },
    "detailedDescription": { ...
```

“Person” schema at schema.org

Person

Canonical URL: http://schema.org/Person

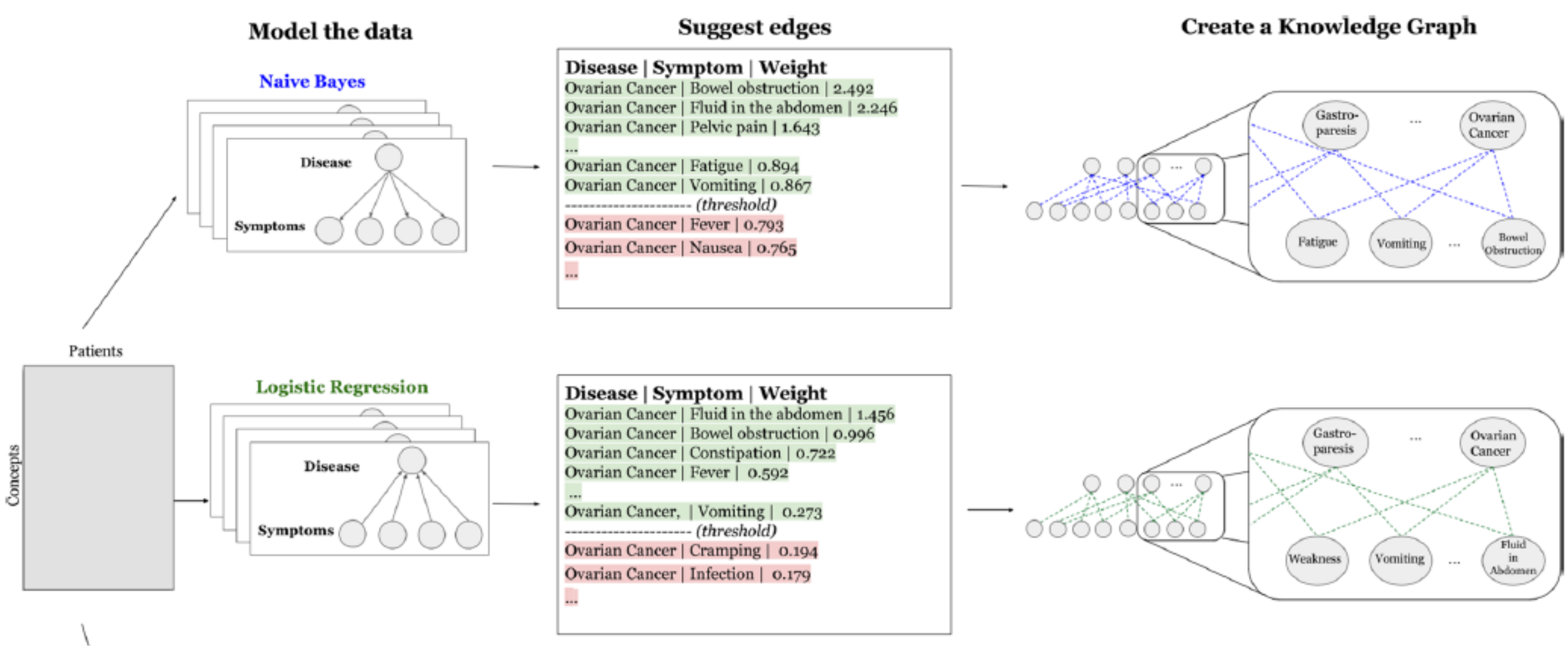
Thing > **Person**
A person (alive, dead, undead, or fictional).

Usage: Over 1,000,000 domains

Property	Expected Type	Description
Properties from Person		
additionalName	Text	An additional name for a Person, can be used for a middle name.
address	PostalAddress or Text	Physical address of the item.
affiliation	Organization	An organization that this person is affiliated with. For example, a school/university, a club, or a team.
alumniOf	EducationalOrganization or Organization	An organization that the person is an alumni of. Inverse property: alumni.
award	Text	An award won by or for this item. Supersedes awards.
birthDate	Date	Date of birth.
birthPlace	Place	The place where the person was born.
brand	Brand or Organization	The brand(s) associated with a product or service, or the brand(s) maintained by an organization or business person.
children	Person	A child of the person.
colleague	Person or URL	A colleague of the person. Supersedes colleagues.
contactPoint	ContactPoint	A contact point for a person or organization. Supersedes contactPoints.
deathDate	Date	Date of death.
deathPlace	Place	The place where the person died.
duns	Text	The Dun & Bradstreet DUNS number for identifying an organization or business person.
email	Text	Email address.
familyName	Text	Family name. In the U.S., the last name of an Person. This can be used along with givenName instead of the name property.
faxNumber	Text	The fax number.
follows	Person	The most generic uni-directional social relation.

Healthcare Knowledge Graph

Maya Rotmensch, Yoni Halpern, Abdulhakim Tlimat, Steven Horng & David Sontag
Learning a Health Knowledge Graph from Electronic Medical Records, Scientific Reports, 2017.
https://www.nature.com/articles/s41598-017-05778-z



The Semantic Web

- Term coined by Tim Berners-Lee
- Common framework for exchange of data across application, enterprise, and community boundaries

- HTML defines how text should look when presented to humans
- Semantic web markup defines how information should be organized to be interpretable by machines

- “Ontology engineer” is a job description now

NELL: Never-Ending Language Learner

Inputs:

- initial ontology
- few examples of each ontology predicate
- the web
- occasional interaction with human trainers

The task:

- run 24x7, forever
- each day:
    1. extract more facts from the web to populate the initial ontology
    2. learn to read (perform #1) better than yesterday

NELL Overview

Running 24x7, January, 12, 2010 – 2019?

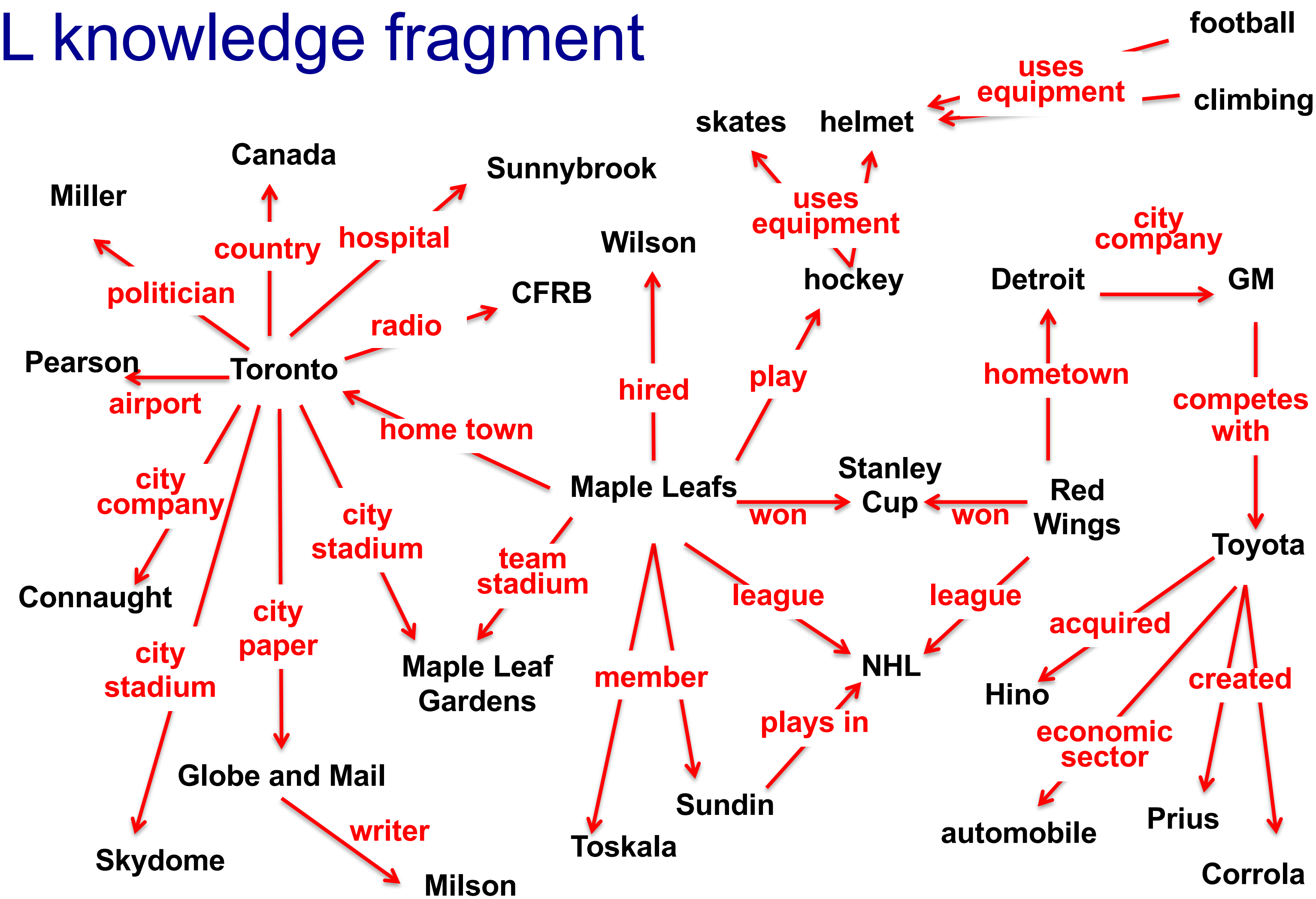
Inputs:

- ontology defining >600 categories and relations
- 10-20 seed examples of each
- 500 million web pages
- 100,000 web search queries per day
- ~ 5 minutes/day of human guidance

Result:

- KB with > 15 million candidate beliefs, growing daily
- learning to reason, as well as read
- automatically extending its ontology

NELL knowledge fragment



NELL Website

- http://rtw.ml.cmu.edu ← follow NELL here
- eg. "diabetes", "Avandia", ,"tea", "IBM", "love" "baseball" "BacteriaCausesCondition" …

Default Approach

Its underconstrained!!

Extract cities:

Paris
Pittsburgh
Seattle
Cupertino

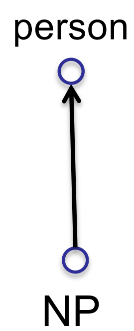
San Francisco
Austin
denial

anxiety
selfishness
Berlin

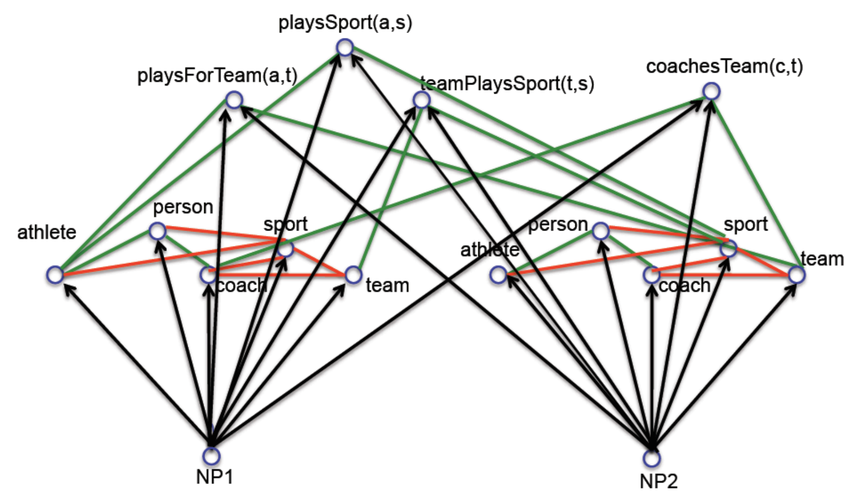
mayor of arg1
live in arg1

arg1 is home of
traits such as arg1

Key Idea 1: Coupled semi-supervised training of many functions



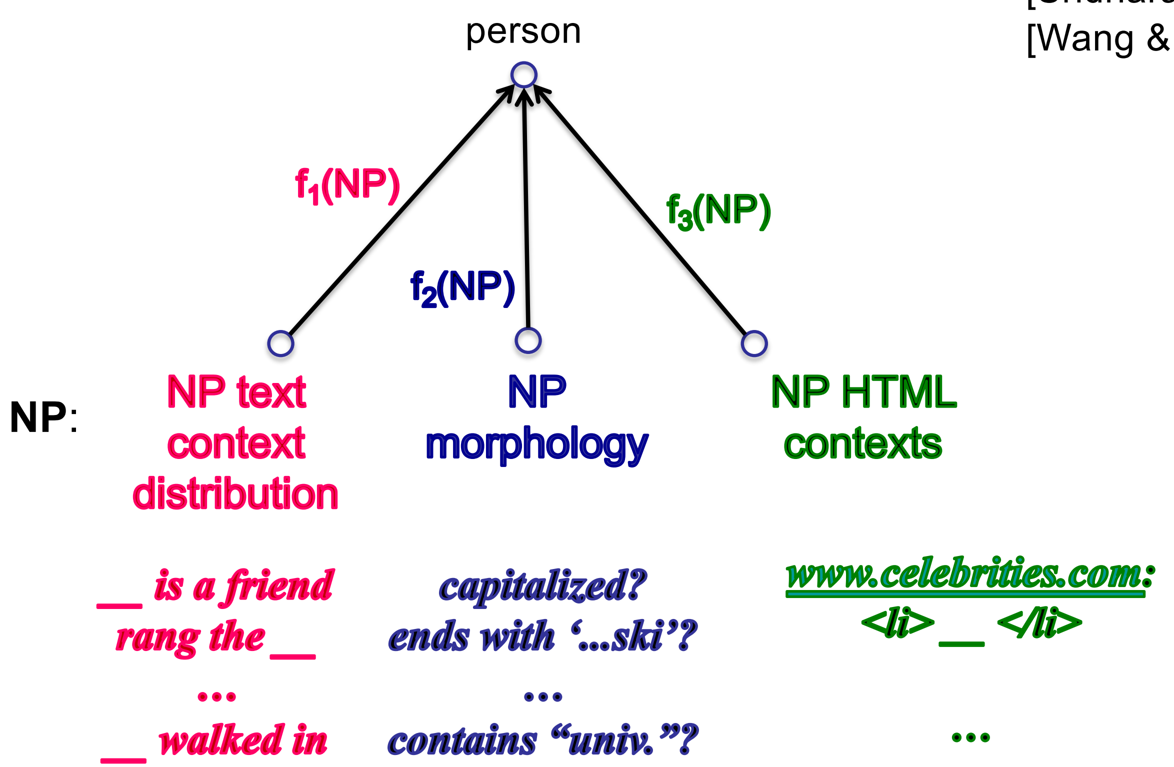
hard
(underconstrained)
semi-supervised
learning problem



much easier (more constrained)
semi-supervised learning problem

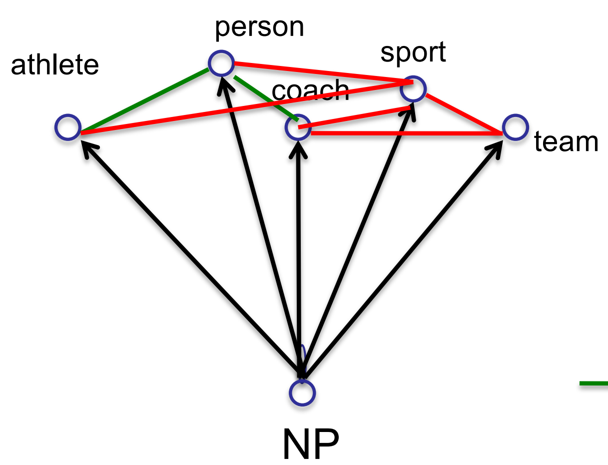
Type 1 Coupling: Co-Training, Multi-View Learning

[Blum & Mitchell; 98]
[Dasgupta et al; 01]
[Ganchev et al., 08]
[Sridharan & Kakade, 08]
[Wang & Zhou, ICML10]



Type 2 Coupling: Multi-task, Structured Outputs

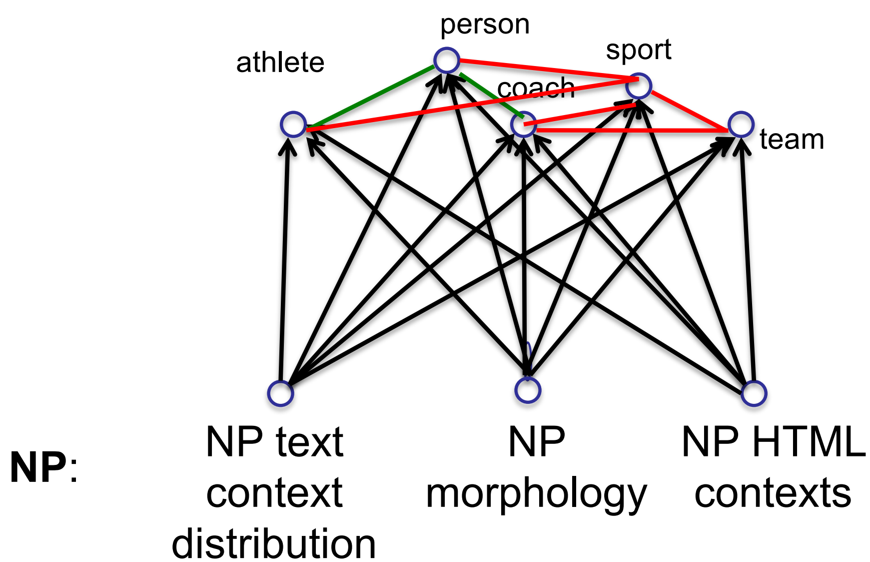
[Daume, 2008]
[Bakhir et al., eds. 2007]
[Roth et al., 2008]
[Taskar et al., 2009]
[Carlson et al., 2009]



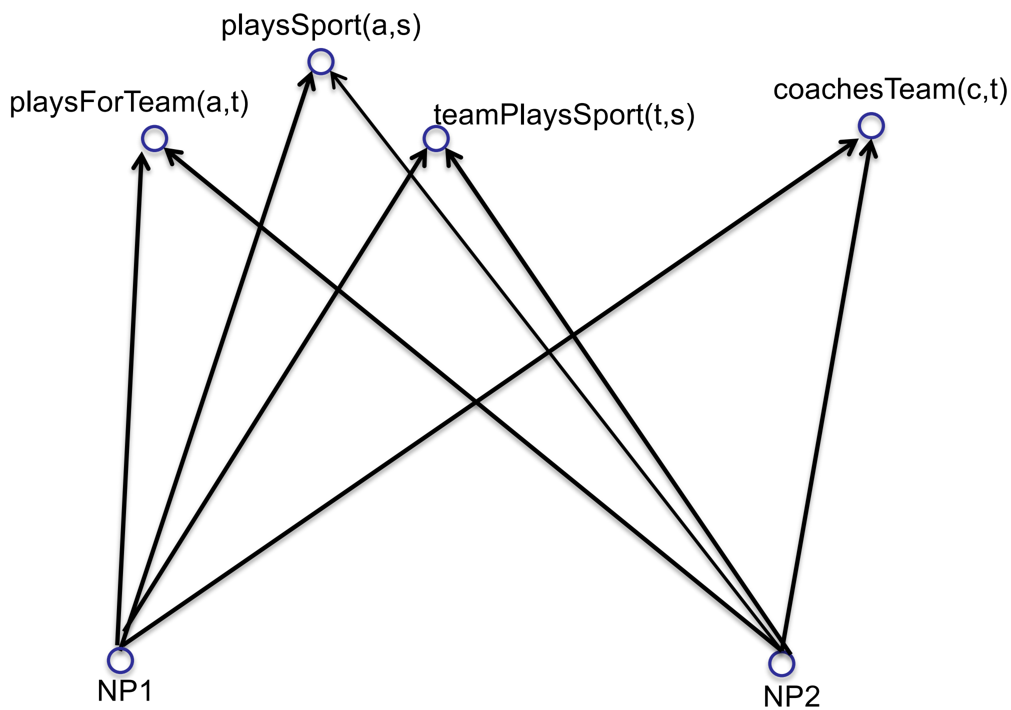
athlete(NP) → person(NP)

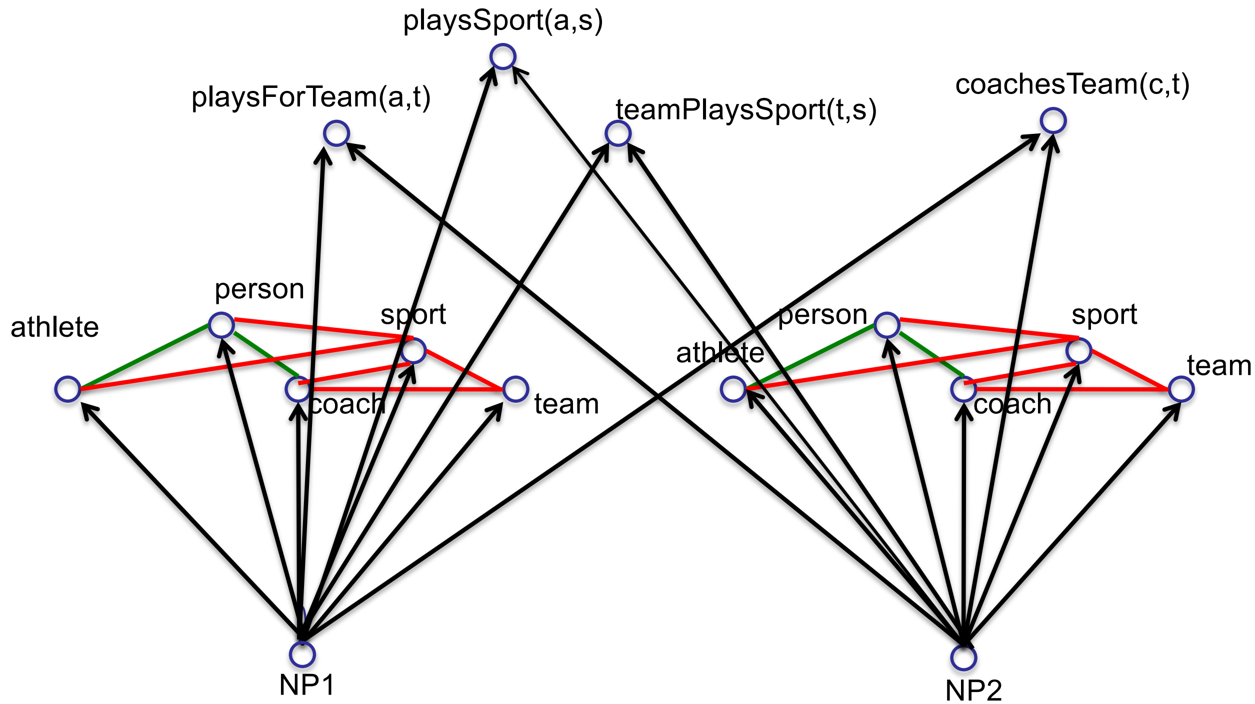
athlete(NP) → NOT sport(NP)
NOT athlete(NP) ← sport(NP)

Multi-view, Multi-Task Coupling



Learning Relations between NP's

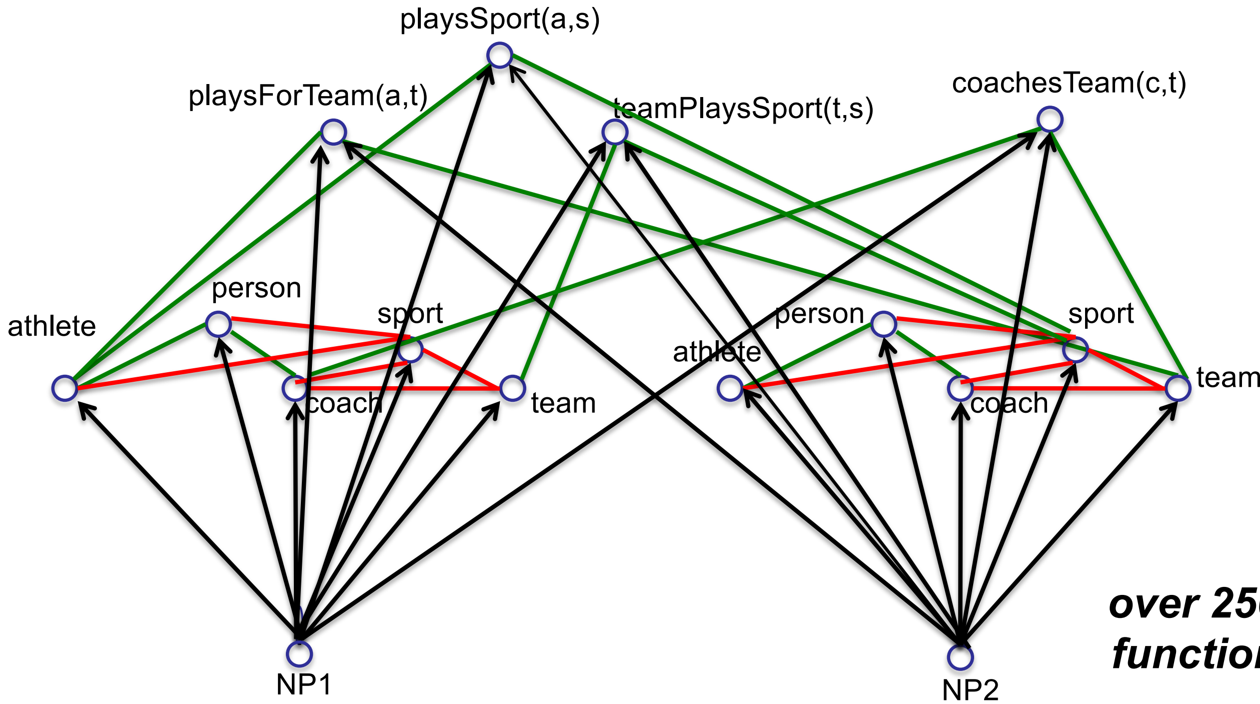




playsSport(a,s)
playsForTeam(a,t)
teamPlaysSport(t,s)
coachesTeam(c,t)
athlete
person
sport
coach
team
NP1
person
sport
athlete
team
coach
NP2

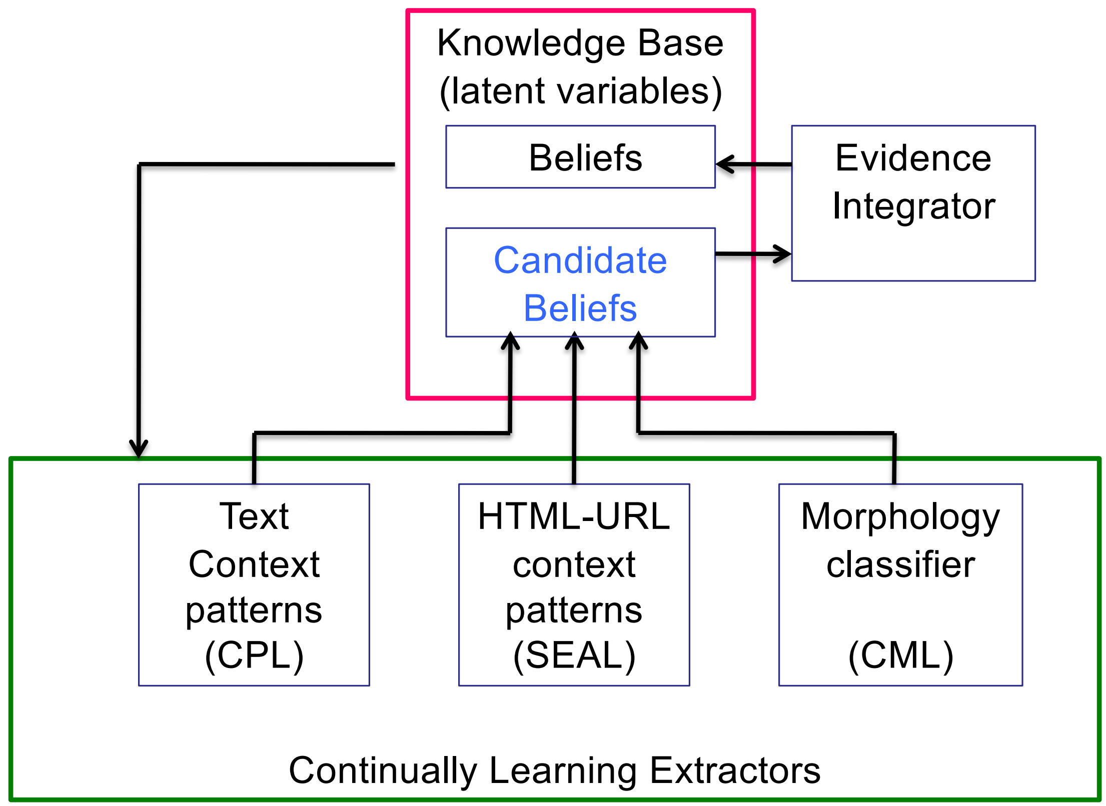
Type 3 Coupling: Argument Types

playsSport(NP1,NP2) → athlete(NP1), sport(NP2)



over 2500 coupled functions in NELL

Basic NELL Architecture



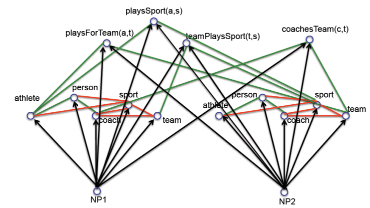
NELL: Learned reading strategies

Plays_Sport(arg1,arg2):

arg1_was_playing_arg2 arg2_megastar_arg1 arg2_icons_arg1
arg2_player_named_arg1 arg2_prodigy_arg1
arg1_is_the_tiger_woods_of_arg2 arg2_career_of_arg1
arg2_greats_as_arg1 arg1_plays_arg2 arg2_player_is_arg1
arg2_legends_arg1 arg1_announced_his_retirement_from_arg2
arg2_operations_chief_arg1 arg2_player_like_arg1
arg2_and_golfing_personalities_including_arg1 arg2_players_like_arg1
arg2_greats_like_arg1 arg2_players_are_steffi_graf_and_arg1
arg2_great_arg1 arg2_champ_arg1 arg2_greats_such_as_arg1
arg2_professionals_such_as_arg1 arg2_hit_by_arg1 arg2_greats_arg1
arg2_icon_arg1 arg2_stars_like_arg1 arg2_pros_like_arg1
arg1_retires_from_arg2 arg2_phenom_arg1 arg2_lesson_from_arg1
arg2_architects_robert_trent_jones_and_arg1 arg2_sensation_arg1
arg2_pros_arg1 arg2_stars_venus_and_arg1 arg2_hall_of_famer_arg1
arg2_superstar_arg1 arg2_legend_arg1 arg2_legends_such_as_arg1
arg2_players_is_arg1 arg2_pro_arg1 arg2_player_was_arg1
arg2_god_arg1 arg2_idol_arg1 arg1_was_born_to_play_arg2
arg2_star_arg1 arg2_hero_arg1 arg2_players_are_arg1
arg1_retired_from_professional_arg2 arg2_legends_as_arg1
arg2_autographed_by_arg1 arg2_champion_arg1 …

If coupled learning is the key,
how can we get new coupling constraints?

Key Idea 2:



Discover New Coupling Constraints

- first order, probabilistic horn clause constraints:

0.93	athletePlaysSport(?x,?y) ←	athletePlaysForTeam(?x,?z)
		teamPlaysSport(?z,?y)

  – connects previously uncoupled relation predicates

  – infers new beliefs for KB

Example Learned Horn Clauses

0.95 athletePlaysSport(?x,basketball) ← athleteInLeague(?x,NBA)

0.93 athletePlaysSport(?x,?y) ← athletePlaysForTeam(?x,?z)
teamPlaysSport(?z,?y)

0.91 teamPlaysInLeague(?x,NHL) ← teamWonTrophy(?x,Stanley_Cup)

0.90 athleteInLeague(?x,?y) ←athletePlaysForTeam(?x,?z),
teamPlaysInLeague(?z,?y)

0.88 cityInState(?x,?y) ← cityCapitalOfState(?x,?y), cityInCountry(?y,USA)

0.62* newspaperInCity(?x,New_York) ← companyEconomicSector(?x,media)
generalizations(?x,blog)

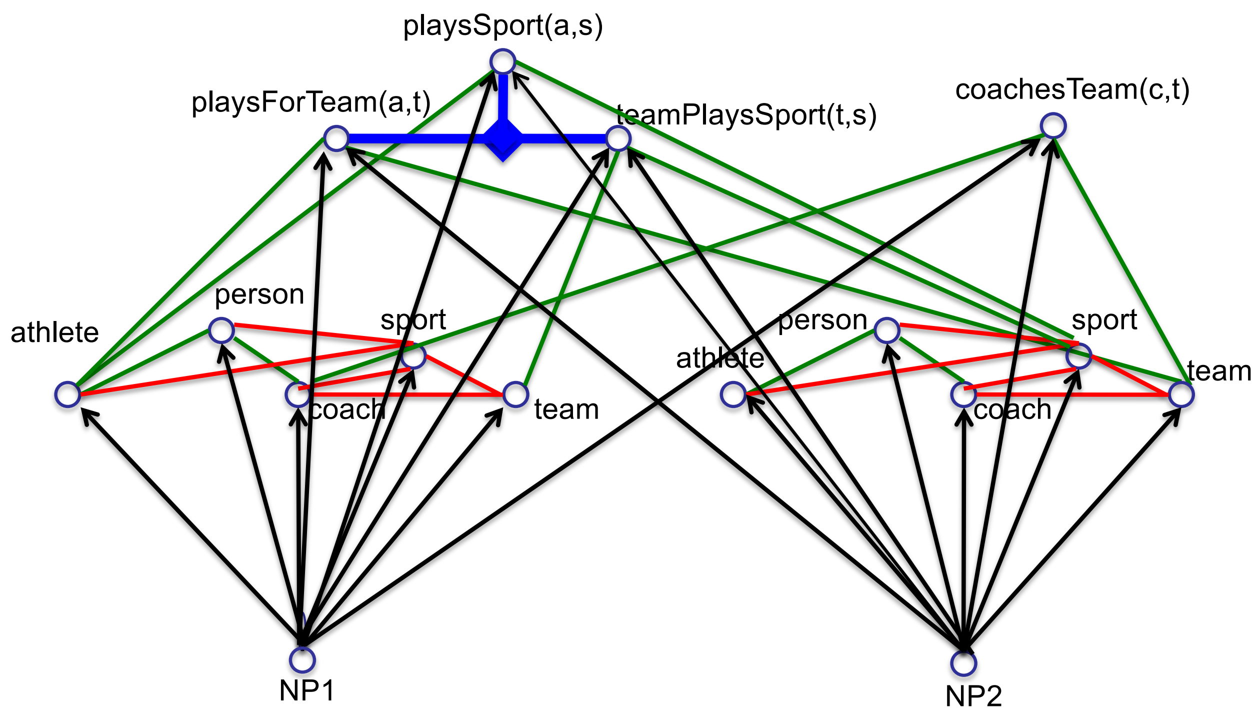
Some rejected learned rules

cityCapitalOfState{?x ?y} ← cityLocatedInState{?x ?y},
teamPlaysInLeague{?y nba}

teamplayssport{?x, basketball} ← generalizations{?x, university}

Learned Probabilistic Horn Clause Rules

0.93 playsSport(?x,?y) ← playsForTeam(?x,?z), teamPlaysSport(?z,?y)



Key Idea 3:

Automatically extend ontology

Ontology Extension (1)

[Mohamed et al., *EMNLP* 2011]

Goal:

- Add new relations to ontology

Approach:

- For each pair of categories C1, C2,
    - co-cluster pairs of known instances, and text contexts that connect them

Example Discovered Relations

[Mohamed et al. *EMNLP* 2011]

Category Pair	Text contexts	Extracted Instances	Suggested Name
MusicInstrument Musician	ARG1 master ARG2 ARG1 virtuoso ARG2 ARG1 legend ARG2 ARG2 plays ARG1	sitar , George Harrison tenor sax, Stan Getz trombone, Tommy Dorsey vibes, Lionel Hampton	Master
Disease Disease	ARG1 is due to ARG2 ARG1 is caused by ARG2	pinched nerve, herniated disk tennis elbow, tendonitis blepharospasm, dystonia	IsDueTo
CellType Chemical	ARG1 that release ARG2 ARG2 releasing ARG1	epithelial cells, surfactant neurons, serotonin mast cells, histomine	ThatRelease
Mammals Plant	ARG1 eat ARG2 ARG2 eating ARG1	koala bears, eucalyptus sheep, grasses goats, saplings	Eat
River City	ARG1 in heart of ARG2 ARG1 which flows through ARG2	Seine, Paris Nile, Cairo Tiber river, Rome	InHeartOf

NELL: recently self-added relations

- athleteWonAward
- animalEatsFood
- languageTaughtInCity
- clothingMadeFromPlant
- beverageServedWithFood
- fishServedWithFood
- athleteBeatAthlete
- athleteInjuredBodyPart
- arthropodFeedsOnInsect
- animalEatsVegetable
- plantRepresentsEmotion
- foodDecreasesRiskOfDisease
- clothingGoesWithClothing
- bacteriaCausesPhysCondition
- buildingMadeOfMaterial
- emotionAssociatedWithDisease
- foodCanCauseDisease
- agriculturalProductAttractsInsect
- arteryArisesFromArtery
- countryHasSportsFans
- bakedGoodServedWithBeverage
- beverageContainsProtein
- animalCanDevelopDisease
- beverageMadeFromBeverage

Key Idea 4: Cumulative, Staged Learning

Learning X improves ability to learn Y

1. Classify noun phrases (NP's) by category
2. Classify NP pairs by relation
3. Discover rules to predict new relation instances
4. Learn which NP's (co)refer to which concepts
5. Discover new relations to extend ontology
6. Learn to infer relation instances via targeted random walks
7. Learn to assign temporal scope to beliefs
8. Learn to microread single sentences
9. Vision: co-train text and visual object recognition
10. Goal-driven reading: predict, then read to corroborate/correct
11. Make NELL a conversational agent on Twitter