# The Battle for the Future of Data Mining

Oren Etzioni, CEO
Allen Institute for AI (AI2)

ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

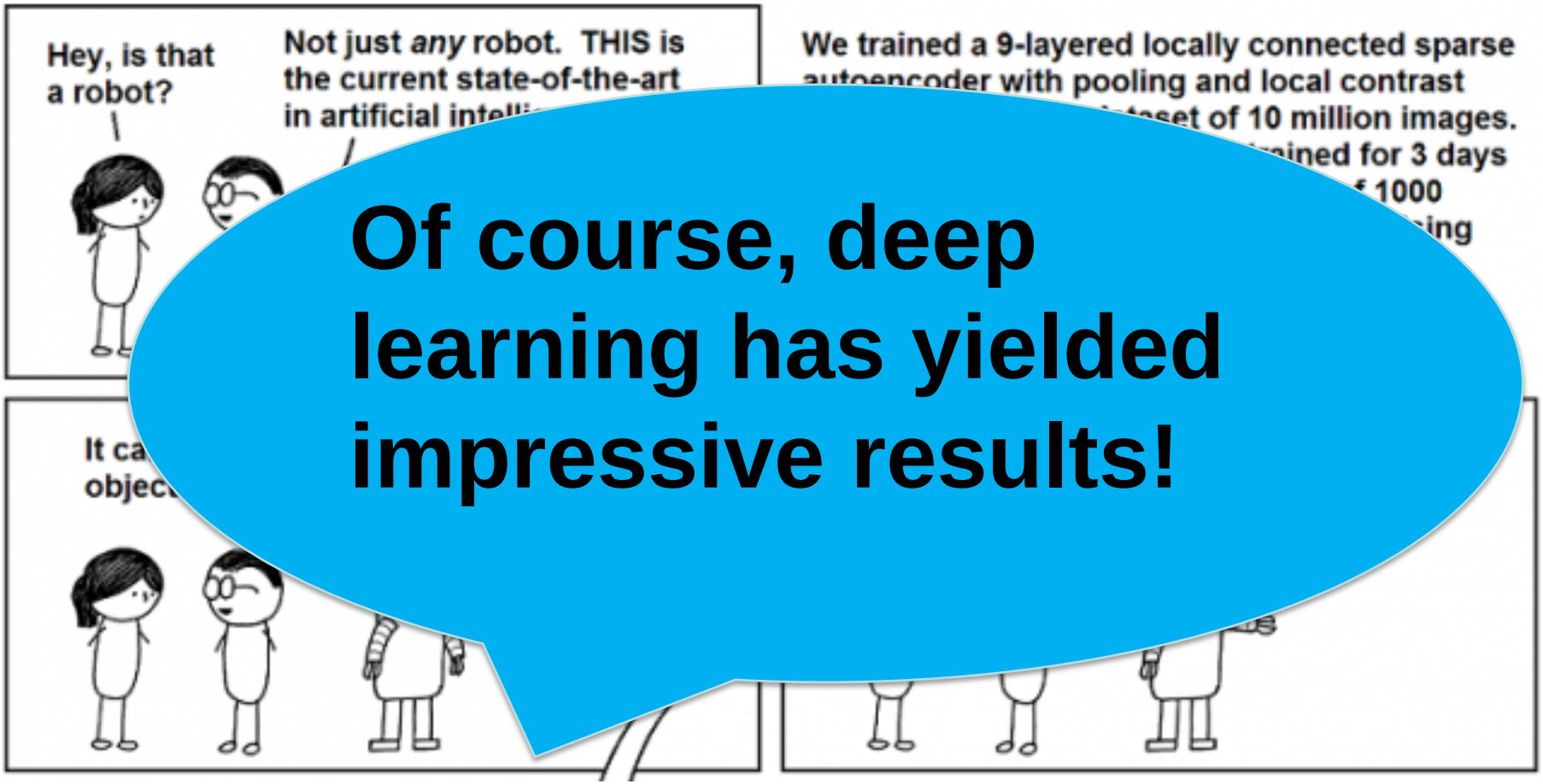
# Big Data Tidal Wave

AI2 ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Deep Learning

ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Some Skepticism about Deep Learning

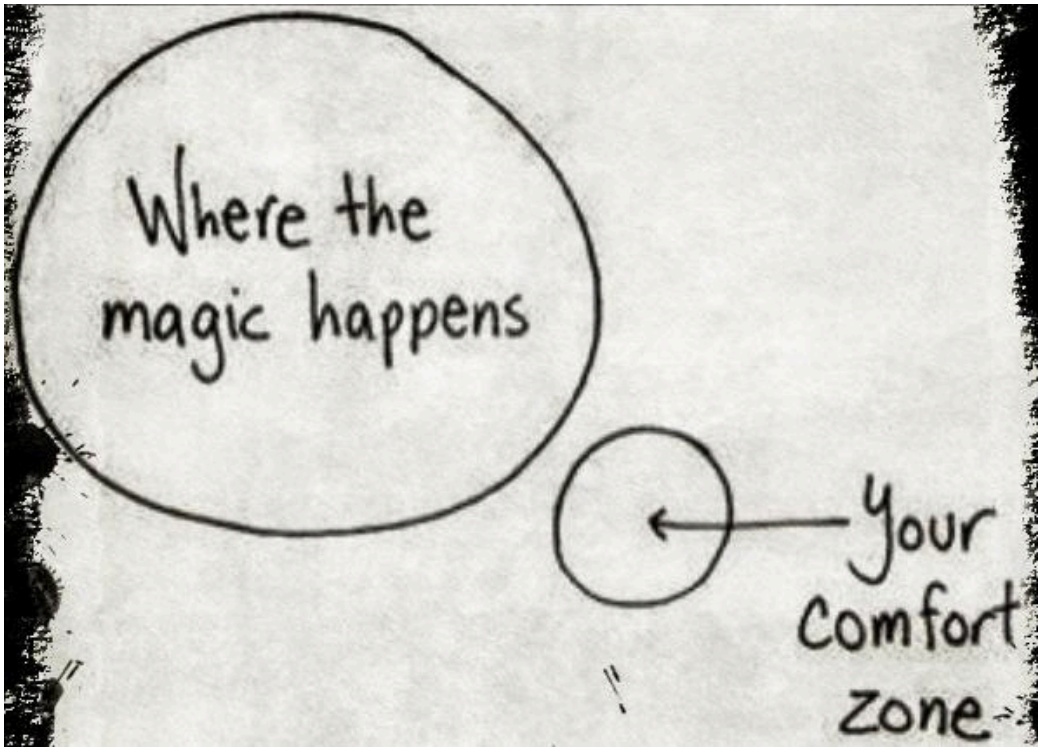





AI2 ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Outline

I. Limits of current paradigm
II. Surprising poverty of Today's Mined KBs
III. Machine Reading
IV. Allen Institute for AI (AI2)






ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# What are the limits of Classification and prediction?

*Note: not arguing for "generative models" over discriminative methods*



ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Example of State of the Art

t Retweets - NYTimes.com - Google Chrome

mes.com/2014/07/03/upshot/a-25-question-twitter-quiz-to-predict-retweets.html

ites

HOME SEARCH

The New York Times

Sendhil Mullainathan

@m_sendhil

Consider these two tweets by Al Gore, both promoting the same article:

Al Gore

@algore

Follow

40% of smartphone users connect to Internet immediately upon awakening, before leaving bed. #TheFuture bit.ly/WYRz39 @TheAtlantic

8:45 AM - 26 Jan 2013

[redacted]

**Based on Tan, Lee, Pang (ACL '14)**

**What about creating Tweets in the first place?**

Can you guess wh… …s retweeted more often?

# Product Recomm

I can't be reduced to a vector of surface variables and correlations!

**But is Netflix lying to us?!**

Prototype in the cloud, then easily move back on-premises
with IBM PureApplication® Solutions
IBM

MUST READ: *Microsoft's Windows 9: Much ado about little, given cloud shift*

*Topic: Big Data*

*Follow via:*

# After a year of medical school, IBM's Watson passes first milestone

***Summary:*** *IBM's year-long residency at Memorial Sloan-Kettering Cancer Center and Wellpoint is finally producing cognitive computing breakthroughs (and two new products).*

By Andrew Nusca for Between the Lines | February 8, 2013 -- 16:45 GMT (08:45 PST)

Follow @editorialiste  *Get the ZDNet Big Data newsletter now*

IBM's Watson has made its first commercially-developed cognitive computing breakthroughs, the company announced this morning. They arrive after more than a year of training with healthcare partners WellPoint and Memorial Sloan-Kettering Cancer Center.

cirba
Driving automation and incredible cost efficiency
Learn how »

*Six Windows 8.1 tips for maximum performance, reliability, and security*

# More Challenges for Classification

- **Learning Arithmetics** = set of crisp rules & procedures

  Distribution insensitive; 100% accuracy

- **Playing Chess:** move choice requires search

  **Deep Blue;** human grand masters

- **Language Understanding**

  Complex grammar

  **Background knowledge**



ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Fundamental Claim

Classification, prediction limited

("structured prediction"; "multi-task learning"
are steps forward)

**Need a process that's constructive, multi-layered, Knowledge-intensive**

**Data → knowledge → theories**
**Theories → reasoning**

# The Bottom Line

**“You can’t play 20 questions with nature and Win.”**

*Allen Newell, 1973*

**Key AI tasks cannot be reduced to classification problems!**

**(using “reasonable” features)**

# Winograd Schema Challenge (Levesque, 2011)

The large ball [illegible] the table
becau[illegible]

**World knowledge is necessary for elementary understanding, but is our technology ready?**

# Just the Facts…

How much do ***you*** know about apples?

- Edible fruit
- Has peel, stem, flesh
- Consists of mostly water
- Ripens on a tree

**at least 20 non-trivial facts + inheritance…**

**Mined KBs are knowledge poor!**



ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# 2014 Knowledge Tour

Google apple +Oren

Web News Images Shopping Maps More Search tools

About 597,000,000 results (0.37 seconds)

**Apple**
www.apple.com/ Apple Inc.
**Apple** designs and creates iPod and iTunes, Mac laptop and desktop computers, the OS X operating system, and the revolutionary iPhone and iPad.
3.8 ★★★★☆ 46 Google reviews · Write a review

2656 NE University Village St, Seattle, WA 98105
(206) 892-0433

**Apple Store**
Shop iPad - Shop iPhone - Shop Mac - Refurbished & Clearance

**iPad**
Introducing iPad Air and the iPad mini with Retina display. Watch ...

**iPhone**
Discover everything iPhone, including the most advanced ...

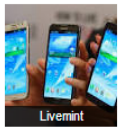
**iTunes**
iTunes is the world's best way to play — and add to — your ...

**Support**
Apple Support is your starting point for help with Apple products ...

**University Village**
The Apple Store is located in University Village, at 2656 NE ...

More results from apple.com »

News for **apple**

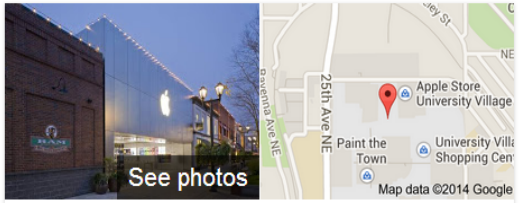



Another **Apple**-Samsung Skirmish Heads to Court
ABC News - 4 minutes ago
The fiercest rivalry in the world of smartphones is heading back to court this week in the heart of the Silicon Valley, with **Apple** and Samsung ...

**Apple** Loop: iPhone 6 Screens, Microsoft Office For iPad, iBeacons For MLB, Secre...
Forbes - by Ewan Spence - 7 hours ago

**Apple** plans to create 'transparent' iPhones with rear camera technology
The Independent - 4 hours ago

More news for **apple**




## Apple Store University Village

Directions Write a review

**Address:** 2656 NE University Village St, Seattle, WA 98105
**Phone:** (206) 892-0433
**Prices:** $$$$
**Hours:** Open today · 11:00 am – 6:00 pm

Reviews
3.8 ★★★★☆ 46 Google reviews

**More reviews:** judysbook.com, yahoo.com, insiderpages.com, servicearea.com

People also search for

University Village Shopping Center | The Mac Store | Apple Store Alderwood Mall | Banana Republic | Best Buy

Are you the business owner? Feedback

See results about

Apple Inc.
Consumer electronics company
Apple Inc. is an American multinational corporation headquartered in Cupertino, California, that ...

"apple fruit" - Google Search - Google Chrome

https://www.google.com/search?q="apple+fruit"

Apps | O-1 Person of Extrao... | WeeklyPlans/Decem... | British Airways - Wh... | 74 | Vulcan - Project Hal... | Seattle Department ... | New Techniques fro... | Extraction Browser - ... | deep learning tutorial | Facebook launches ... | An AI that mimics o...

Google "apple fruit" +Oren

Web Images Shopping Videos News More Search tools

About 365,000 results (0.27 seconds)

**Apple - Wikipedia, the free encyclopedia**
en.wikipedia.org/wiki/Apple ▾ Wikipedia ▾
The apple is the pomaceous fruit of the apple tree, species Malus domestica in the rose family (Rosaceae). It is one of the most widely cultivated tree fruits, and ...
Apple Inc. - Disambiguation - Malus sieversii - List of apple cultivars

**Apple fruit** nutrition facts and health benefits - Nutrition and ...

www.nutrition-and-you.com/**apple-fruit**.html ▾
by Umesh Rudrappa - in 70 Google+ circles
Delicious and crunchy **apple fruit** is one of the popular fruits containing an impressive list of antioxidants and essential nutrients required for good health...

**apple (fruit** and tree) -- Encyclopedia Britannica
www.britannica.com/EBchecked/topic/.../apple ▾ Encyclopaedia Britannica ▾
Jul 19, 2013 - Malus domestica fruit of the domesticated tree Malus domestica (family Rosaceae), one of the most widely cultivated tree fruits. The apple is a ...

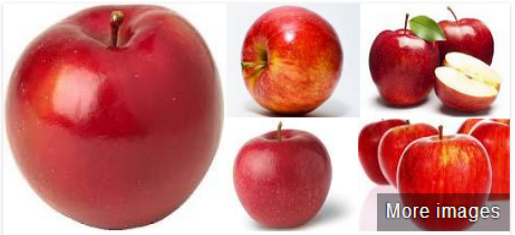
Cashew-**Apple Fruit** Growing in the Florida Home Landscape...
edis.ifas.ufl.edu/hs377 ▾ Institute of Food and Agricultural Sciences ▾
Cashew-**Apple Fruit** Growing in the Florida Home Landscape1. John McLaughlin, Carlos Balerdi and Jonathan Crane2. Scientific Name: Anacardium ...

The Apple - "King" of the Fruit - Nature's Wonderland - Jewish...
www.chabad.org/kids/article_cdo/aid/114763/.../Apples.htm ▾ Chabad ▾
History of the Apple. In honor of Tu-B'Shevat (15th day of Sh'vat), we will now talk about a fruit - a very special fruit - the Apple, "King" of all fruits (in Hebrew it is ...

Sneaky Green **Apple Fruit** Leather Recipe - Oh My Veggies

ohmyveggies.com/recipe-sneaky-green-**apple-fruit**-leather/ ▾
6 hrs 10 mins
Aug 15, 2013 - This sugar-free **apple fruit** leather recipe is hiding a secret--it's made with spinach! But it tastes so good, you'll never know it's in there.

Truffle - Edible Arrangements




## Apple

Fruit

The apple is the pomaceous fruit of the apple tree, species Malus domestica in the rose family. It is one of the most widely cultivated tree fruits, and the most widely known of the many members of genus Malus that are used by humans. Wikipedia

### Nutrition Facts

Apple ▾

**Amount Per** 1 medium (3" dia) (182 g) ▾

**Calories** 95

| | % Daily Value* |
|---|---|
| **Total Fat** 0.3 g | 0% |
| Saturated fat 0.1 g | 0% |
| Polyunsaturated fat 0.1 g | |
| Monounsaturated fat 0 g | |
| **Cholesterol** 0 mg | 0% |
| **Sodium** 2 mg | 0% |
| **Potassium** 195 mg | 5% |
| **Total Carbohydrate** 25 g | 8% |
| Dietary fiber 4.4 g | 17% |
| Sugar 19 g | |
| **Protein** 0.5 g | 1% |

| | | | |
|---|---|---|---|
| Vitamin A | 1% | Vitamin C | 14% |
| Calcium | 1% | Iron | 1% |
| Vitamin D | 0% | Vitamin B-6 | 5% |
| Vitamin B-12 | 0% | Magnesium | 2% |

apple fruit - Wolfram|Alpha - Google Chrome
https://www.wolframalpha.com/input/?i=apple+fruit
Apps | O-1 Person of Extrao... | WeeklyPlans/Decem... | British Airways - Wh... | 74 | Vulcan - Project Hal... | Seattle Department ... | New Techniques fro... | Extraction Browser - ... | deep learning tutorial | Facebook launches ... | An AI that mimics o...
PRO MOBILE APPS PRODUCTS EXAMPLES BLOG ABOUT WOLFRAM SITES

*Making Wolfram|Alpha free to all wouldn't be possible without the support of our Pro subscribers.*
Help us make even more of the world computable in 2014 by going Pro. Go Pro

Share:
more

WolframAlpha computational... knowledge engine

apple fruit

Examples Random

Assuming "apple" is a food | Use as a general material instead

Assuming apple | Use prepared apples or more ▾ instead

Assuming any type of apple | Use apple, with skin or apple, without skin instead

Assuming any type of fruit | Use abiyuch, generic or more ▾ instead

Input interpretation:

| apple | amount | 1 apple | fruit | amount | 140 grams |
|---|---|---|---|---|---|

Individual nutrition facts: Total nutrition facts

Apple:

| serving size 1 apple (182 g) | |
|---|---|
| **total calories** 91 | fat calories 2 |
| | % daily value* |
| **total fat** 273 mg | **0%** |
| saturated fat 45 mg | **0%** |
| trans fat | |
| **cholesterol** 0 g | **0%** |
| **sodium** 910 µg | **0%** |

New to Wolfram|Alpha?

Take the Tour »

Take Wolfram|Alpha anywhere...

Concept: "apple" (Mx8Ngh4rvVipdpwpEbGdrcN5Y29ycB4rvVjBnZwpEbGdrcN5Y29ycA) - Google Chrome

sw.opencyc.org/2012/05/10/concept/en/Apple

Apps | O-1 Person of Extrao... | WeeklyPlans/Decem... | British Airways - Wh... | 74 | Vulcan - Project Hal... | Seattle Department ... | New Techniques fro... | Extraction Browser - ... | deep learning tutorial | Facebook launches ... | An AI that mimics o...

OpenCyc (Current): [ http://sw.opencyc.org/concept/Mx8Ngh4rvVipdpwpEbGdrcN5Y29ycB4rvVjBnZwpEbGdrcN5Y29ycA ]
OpenCyc (Versioned): [ http://sw.opencyc.org/2012/05/10/concept/Mx8Ngh4rvVipdpwpEbGdrcN5Y29ycB4rvVjBnZwpEbGdrcN5Y29ycA ]
OpenCyc (Readable): [ http://sw.opencyc.org/2012/05/10/concept/en/Apple ]

Search

# OpenCyc Collection: apple

**Unique ID: [ Mx8Ngh4rvVipdpwpEbGdrcN5Y29ycB4rvVjBnZwpEbGdrcN5Y29ycA ]**
**English ID: [ (FruitFn AppleTree) ]**
**English Aliases: [ "apples", "fruit of the apple", "fruit of the apple tree", "fruit of the apple trees", "fruit of the apples", "fruit of the Malus pumila" ]**

The collection of individual apples.

**A Type of:** accessory fruit, edible fruit

**Instance of:** type of food, type of life stage, type of object whose instances do not physically overlap

**Subtypes:** Adams Pearmain apple, Ambrosia apple, Braeburn apple, Gala apple, Golden Delicious apple, Granny Smith apple, Jazz apple, McIntosh apple, red delicious apple

**Instances:**

**Same as:**

Copyright © 2001-2012 Cycorp, Inc.

ConceptNet 5 - Google Chrome
conceptnet5.media.mit.edu/web/c/en/apple
Apps O-1 Person of Extrao... WeeklyPlans/Decem... British Airways - Wh... 74 Vulcan - Project Hal... Seattle Department ... New Techniques fro... Extraction Browser - ... deep learning tutorial Facebook launches ... An AI that mimics o...



Get /c/en/apple in JSON format

# apple

apple — *RelatedTo* → fruit
*apple is related to a fruit*

apple — *RelatedTo* → red fruit
*apple is related to red fruit*

apple — *RelatedTo* → core
*apple is related to core*

apple — *Antonym* → orange
*apple is the opposite of orange*

apple — *RelatedTo* → computer
*apple is related to a computer*

apple — *RelatedTo* → green
*apple is related to green*

apple — *RelatedTo* → eve
*apple is related to eve*

apple — *RelatedTo* → mac
*apple is related to mac*

apple — *RelatedTo* → pie
*apple is related to pie*

apple — *HasProperty* → green
*apple can be green*

apple — *RelatedTo* → red
*apple is related to red*

apple — *RelatedTo* → a
*apple is related to a*

apple — *RelatedTo* → tree
*apple is related to tree*

apple — *HasProperty* → red
*apple can be red*

apple — *RelatedTo* → seed
*apple is related to seeds*

apple — *RelatedTo* → round
*apple is related to round*

apple — *RelatedTo* → macintosh
*apple is related to macintosh*

apple — *RelatedTo* → pear
*apple is related to a pear*

apple — *RelatedTo* → orange
*apple is related to oranges*

apple — *ReceivesAction* → eat
*an apple can be eaten*

# Observations from the Tour

Current KBs are surprisingly "knowledge poor"

- Not specific to Apples

**What about obtaining knowledge from text?**



ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# III. Machine Reading

Auto-Text to Knowledge



Source: DARPA, Machine Reading initiative

# Information Extraction (IE)

IE(sentence) = [tuple, confidence]

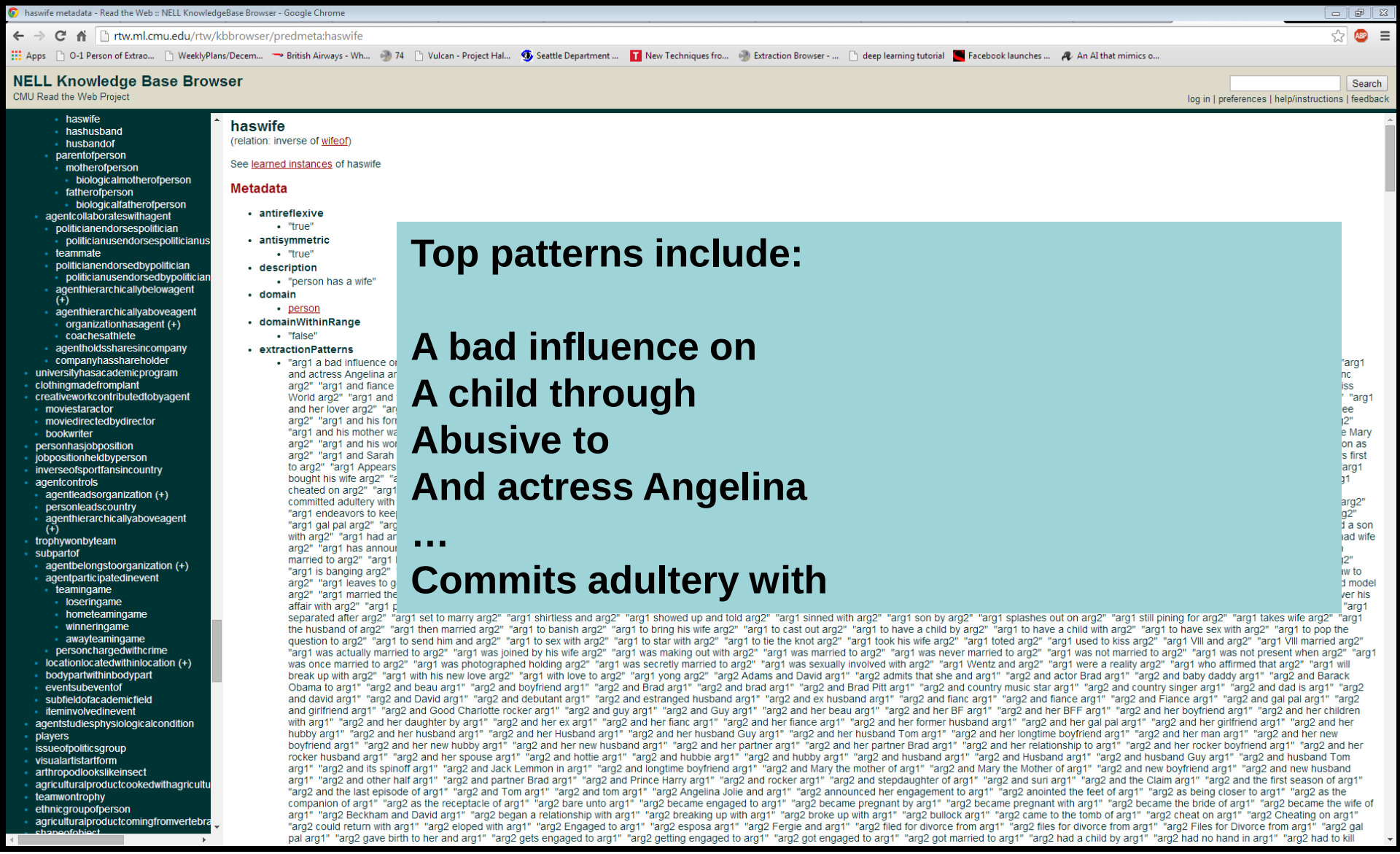
"Edison, by all accounts, was the inventor of the light bulb."
➔ **invented(Edison, light bulb), 0.98**

Typically, IE requires:

- Pre-specified relations
- Hand-labeled training examples
- Lexicalized features/patterns

**Consider Tom Mitchell's NELL System**

# NELL Lexical Patterns for “haswife”






ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

Ask NELL - Read the Web :: NELL KnowledgeBase Browser - Google Chrome

rtw.ml.cmu.edu/rtw/kbbrowser/special:asknell

Apps | 0-1 Person of Extrao... | WeeklyPlans/Decem... | British Airways - Wh... | 74 | Vulcan - Project Hal... | Seattle Department ... | New Techniques fro... | Extraction Browser - ... | deep learning tutorial | Facebook launches ... | An AI that mimics o...

# NELL Knowledge Base Browser

CMU Read the Web Project

log in | p

categories | relations

- everypromotedthing
  - abstractthing
    - creativework
      - book
      - poem
      - lyrics
      - musicalbum
      - musicsong
      - televisionshow
      - movie
      - visualartform
    - species
      - animal
        - vertebrate
          - bird
          - fish
          - reptile
          - mammal
          - amphibian
        - invertebrate
          - arthropod
            - insect
            - crustacean
            - arachnid
          - mollusk
      - archaea
      - bacteria
      - fungus
      - plant
    - physicalaction
    - celltype
    - event
      - eventoutcome
      - conference
        - mlconference
      - sportsevent
        - olympics
        - race
        - grandprix
        - sportsgame
      - convention
      - crimeorcharge
      - filmfestival
      - musicfestival
      - weatherphenomenon
      - election
      - militaryeventtype
        - militaryconflict
    - color
    - date
      - dayofweek
      - year
      - dateliteral
      - month
    - programminglanguage
    - protein
    - chemical
    - buildingmaterial
    - mldataset

## Ask NELL on-demand results:

### 4 possible entities found

Click to change visible entity:

- apple_inc (organization)
- apple (magazine)
- apple (company)
- apple (hotel)

---

### apple_inc (organization)

literal strings: apple computers, Apple inc, -Apple, Apple, Apple Computer Inc., ,Apple, â€œApple, APPLE, Apple Computer, Apple Inc, apple computer, APPLE COMPUTER, APple, apple-computer, apple, Apple computer, Apple computer, Apple Computers

#### categories

- company(100.0%)
  - SEAL @112 (100.0%) on 03-jun-2010 why? using apple
  - CPL @722 (100.0%) on 06-apr-2013 why? using apple
  - Seed
  - MBL @728 (99.9%) on 29-apr-2013 why? using concept:biotechcompany:apple_inc



anewell.JPG

# NELL Knowledge Base

categories

- relatedto
  - visualartmoveme
  - personinacadem
  - generalizationof
    - athletessucha
    - academicfields
    - chemicaltypeh
    - astronautssuc
    - automobileeng
    - airportsuchasa
    - amphibiansuc
    - aquariumssuc
    - arteriessuchas
    - animaltypehas
      - animalsucha
        - animalsuc
        - animalsuc
      - animalsucha
      - inverseofarti
    - professiontype
    - productinstan
    - booksuchasbo
    - vehicletypehas
    - touristattractio
    - celebritysucha
    - archaeasucha

## obama (politician)

literal strings: OBAMA, OBama, Obama, obama

## Help NELL Learn!

**NELL wants to know if these beliefs are correct.**
**If they are or ever were, click thumbs-up. Otherwise, click thumbs-down.**

- obama is a politician
- obama is a male
- obama is a person who belongs to the organization house (governmentorganization)
- obama is a person born on the date n1961 (dateliteral)
- obama is a person who graduated from the university harvard (university)
- obama is a person who graduated from the university harvard_law_school (university)

## categories

- **politician**(100.0%)
  - SEAL @9 (75.0%) on 13-jan-2010 [ 1 2 ] using obama
  - CPL @722 (88.0%) on 06-apr-2013 [ "_ is a good president" "_ defends choice" "_ beats McCain" massive scope" "_ is the only President" "_ is the new Hitler" "_ is half-white" "_ is an empty suit" "_ President" "_ is leading McCain" "McCain trails _" "Kennedy endorsed _" "_ 's Inauguration Day" "_ campaign" "York Times endorses _" "_ 's Inaugural Address" "more popular votes than _" "_ 's his

# *Open* Information Extraction (2007)

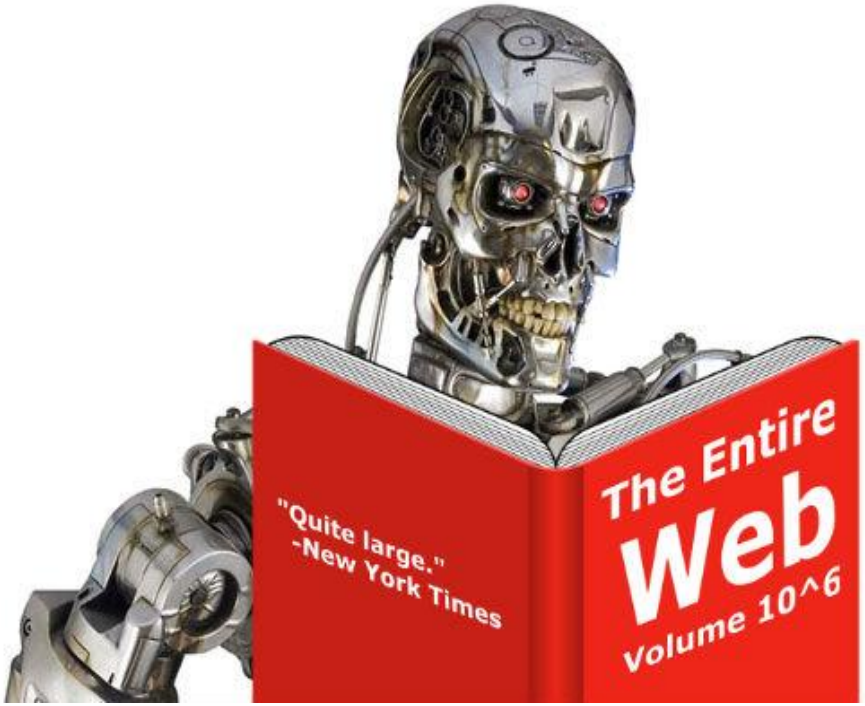
**Question:** can we leverage regularities in language to extract information in a relation-independent way?

**Relations often:**

- anchored in verbs.
- exhibit simple syntactic form

**Virtues:**

- No hand-labeled data
- "No sentence left behind"
- Exploit redundancy of Web



# History of Open IE

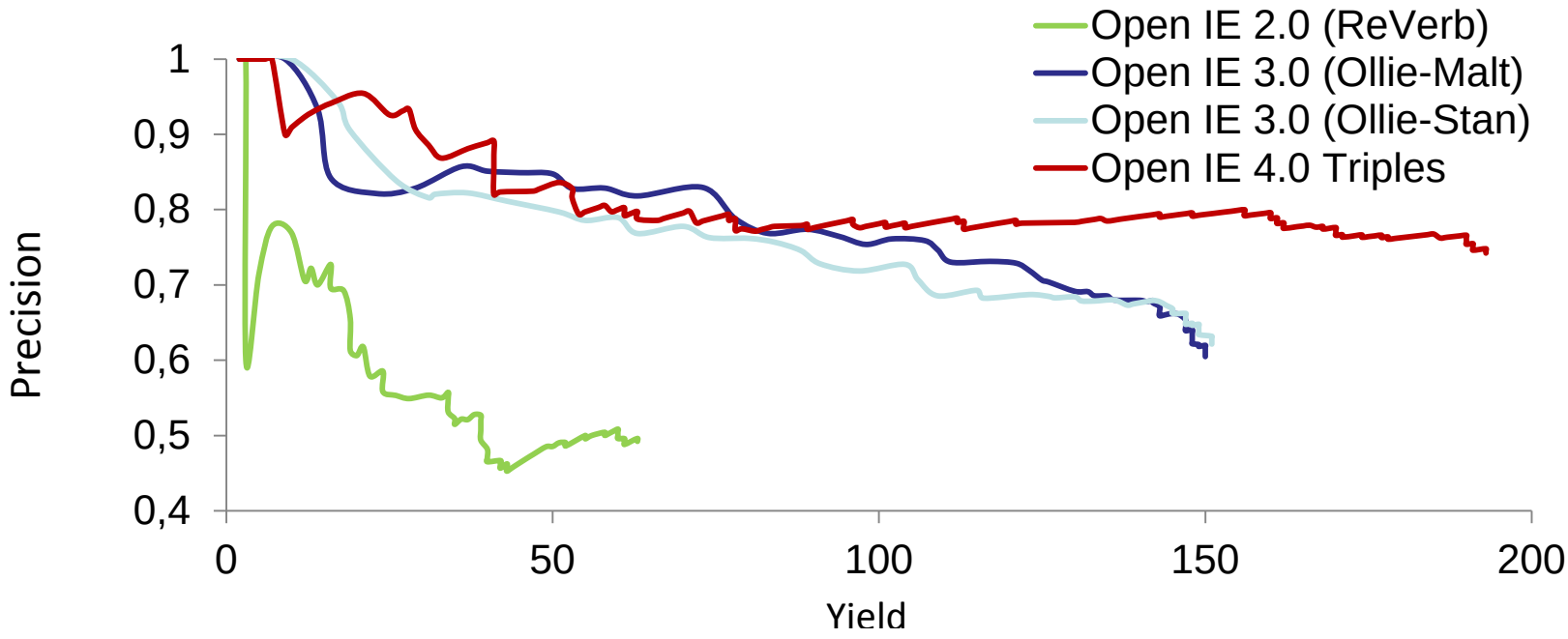
2007: 1.0 (TextRunner, auto-labeled examples, CRF)

2011: 2.0 (ReVerb, simple model of verb-based relations)

2012: 3.0 (Ollie, parser, verbs + nouns)

2013: 4.0 (semantic role labeling, n-ary relations)

*2014: 5.0 (multiple sentences, complex structures)*






ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

Open Information Extraction - Google Chrome

openie.cs.washington.edu/search?arg1=apple&rel=&arg2=&corpora=

Apps | O-1 Person of Extrao... | WeeklyPlans/Decem... | British Airways - Wh... | 74 | Vulcan - Project Hal... | Seattle Department ... | New Techniques fro... | Extraction Browser - ... | deep learning tutorial | Facebook launches ... | An AI that mimics o...

# Open Information Extraction

Turing Center, University of Washington

Argument 1: apple  Relation:  Argument 2:  All  Search

**408 answers** from **11543 sentences** (cached)

Do you mean...

## Apple Inc. (345)

Apple Inc. (NASDAQ: AAPL; formerly Apple Computer, Inc.) is an American multinational corporation that designs and sells consumer electronics, computer software, and personal computers. The company's best-known hardware products are the Macintosh line of computers, the iPod, the iPhone and the iPad. Its software includes the Mac OS X operating system; the iTunes media browser; the iLife suite of multimedia and creativity software; the iWork suite of productivity software; Aperture, a...read more

## Apple (14)

The apple is the pomaceous fruit of the apple tree, species Malus domestica in the rose family (Rosaceae). It is one of the most widely cultivated tree fruits, and the most widely known of the many members of genus Malus that are used by humans. Apples grow on small, deciduous trees. The tree originated in Western Asia, where its wild ancestor, Malus sieversii, is still found today. Apples have been grown for thousands of years in Asia and Europe, and were brought to North America by...read more

## IPad (4)

The iPad ( /ˈaɪpæd/ EYE-pad) is a line of tablet computers designed and marketed by Apple Inc., primarily as a platform for audio-visual media including books, periodicals, movies, music, games, apps and web content. Its size and weight fall between those of contemporary smartphones and laptop computers. The iPad runs on iOS, the same operating system used on Apple's iPod Touch and iPhone, and can run its own applications as well as iPhone applications. Without modification or a developer...read more

## Apple Store (1)

The Apple Retail Store is a chain of retail stores owned and operated by Apple Inc., dealing in computers and consumer electronics. The stores sell Macintosh personal computers, software, iPods, iPads, iPhones, third-party accessories, and other consumer electronics such as Apple TV. Many stores feature a Theatre for presentations and workshops and a Studio for training with Apple products; all stores offer a Genius Bar for technical support and repairs, as well as free workshops available...read more

Open Information Extraction - Google Chrome

openie.cs.washington.edu/results/?arg1=entity%3AApple&justResults=false

Apps | O-1 Person of Extrao... | WeeklyPlans/Decem... | British Airways - Wh... | 74 | Vulcan - Project Hal... | Seattle Department ... | New Techniques fro... | Extraction Browser - ... | deep learning tutorial | Facebook launches ... | An AI that mimics o...

# Open Information Extraction

Turing Center UNIVERSITY OF WASHINGTON

Argument 1: entity:Apple  Relation:  Argument 2:  All  Search

**289 answers** from **2008 sentences** (results truncated)

## Apple

The apple is the pomaceous fruit of the apple tree, species Malus domestica in the rose family (Rosaceae). It is one of the most widely cultivated tree fruits, and the most widely known of the many members of genus Malus that are used by humans. Apples grow on small, deciduous trees. [illegible] Malus sieversii, is still found [illegible]

**Open IE suffers from Attention Deficit Disorder!**

all | ingredient (37) | food (33) | [illegible]

- are tender (263)
- fall to the ground (68)
- are in bloom (68)
- grow on **Tree** (58)
- contain **Pectin** (53)
- are a good source of **Dietary fiber** (40)
- produces **Apple** (34)
- are rich in **Pectin** (33)
- are in season (32)
- are high in **Dietary fiber** (30)
- fall from **Tree** (24)
- is cut in half (22)

# Dr. Tony Fader (AI2): Open Question Answering KDD 2014

ReVerb

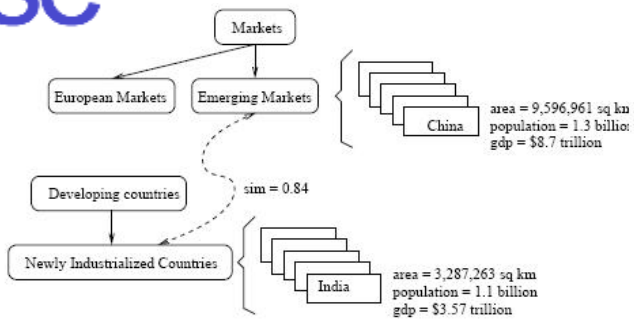
Argument 1:
what/who
Relation:
contains
Argument 2:
antioxidant
all
ingredient (38)
Vegetable (42)
Green tea (23)
Fruit (22)
Wine (20)

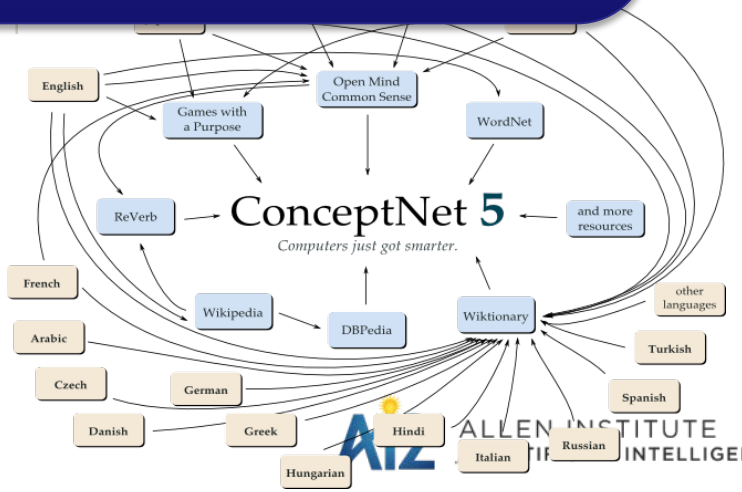
Freebase

Freebase
Find...
Topic
Tao Lin en
mid: /m/025wtg1
Tao Lin is an American
film production company.
Properties
I18n
People /people
Person /people/person
Date of birth /people/person/date_of_birth
7/2/1983
Place of birth /people/person/place_of_birth
Alexandria
Country of nationality /people/person/nationality
United States of America
Gender /people/person/gender
Male
Profession /people/person/profession

Read
Research Proj
Home
NELL: Never-

ProBase




ALLEN INSTITUTE FOR ARTIFICIAL INTELLIGENCE

# IV. The Allen Institute for AI (AI2)

"It is one thing to feed *The Tale of Two Cities* into a computer. It's another to have the computer understand what's being said."

**Paul Allen**, Microcomputer Interface, **1977.**

**AI2 Mission:** *high-impact AI research for the common good*

# Mission + Talent = Impact

| **Time Line** | |
|---|---|
| AI2 launched | Jan. 2014 |
| Team of 30 | August |
| Team of 50;<br>Machine Vision Project | Dec. 2015 |



ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# AI2 Overview (Nonprofit Research Institute)

- Experimental science projects
- Targeted collaborations
- Launched AI Allen Distinguished Investigator (ADI) program
- Open source projects

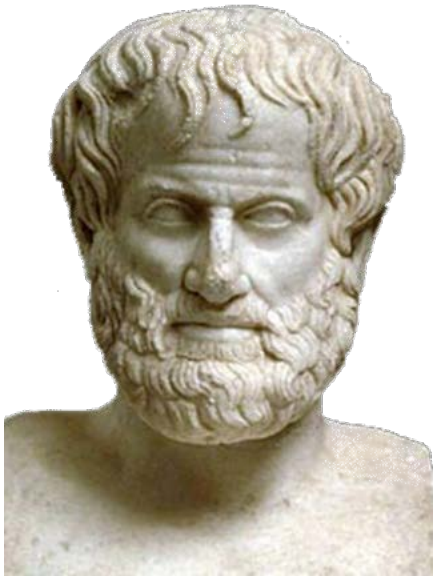
**Small number of “Move the needle” projects**



ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Flagship Project: Aristo

## Project Methodology:



1. Externally-defined challenge tasks
2. Training data + **unseen** test data (“as is”)
3. Measurable progress, clear focus

Key differences with Watson:

1. Deeper semantics & inference
2. Open model: publish, collaborate, open source

# Aristo: 4th, 5th, 6th … Grade Science Tests

THE UNIVERSITY OF THE STATE OF NEW YORK

**GRADE 4**

**ELEMENTARY-LEVEL**

**SCIENCE TEST**

**WRITTEN TEST**

**JUNE 3, 2013**

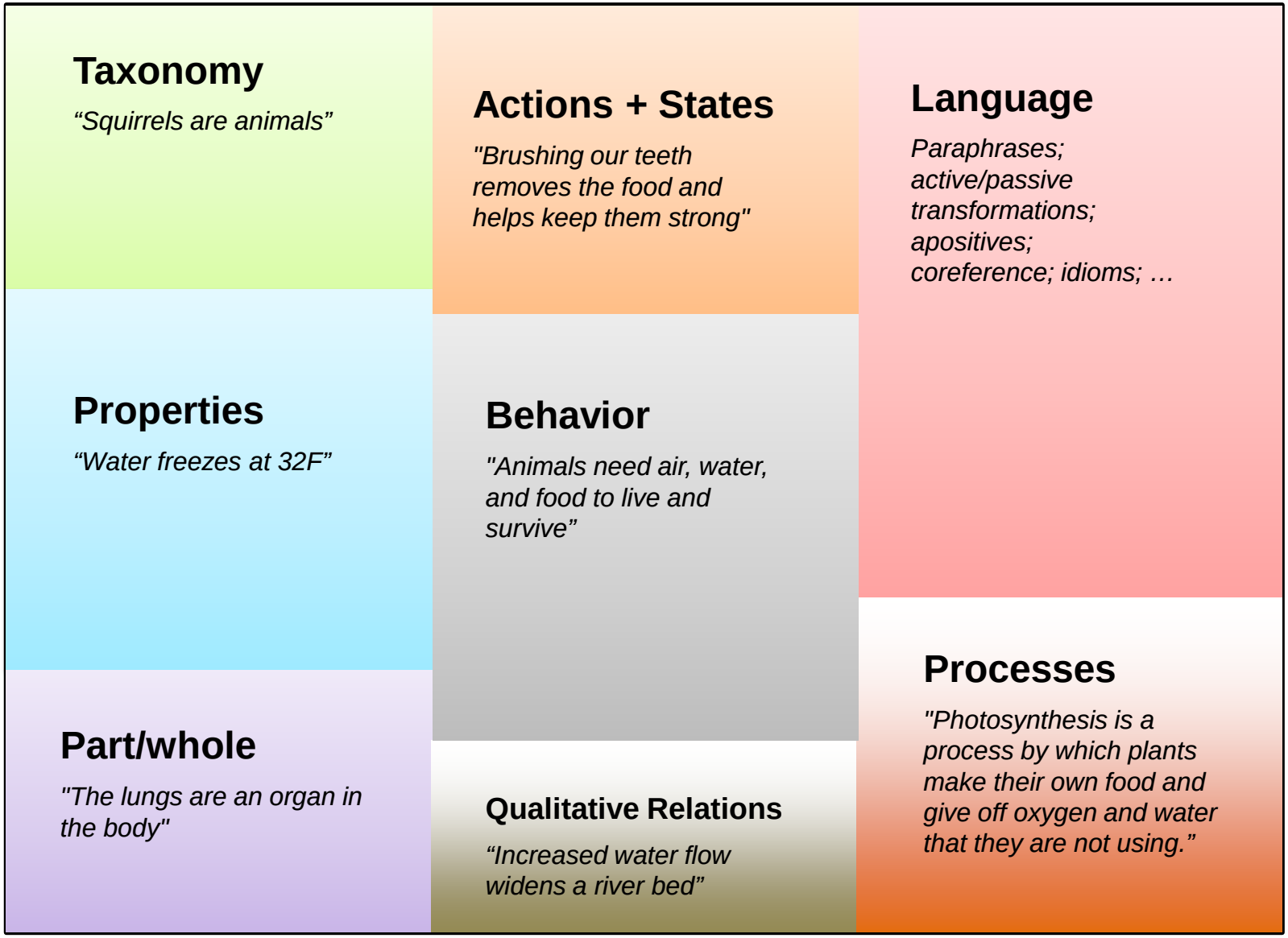
**12** Which force causes a bicycle to slow down when the brakes are used?

A friction
B electricity
C gravity
D magnetism



ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Factual Knowledge for 4th Grade Science

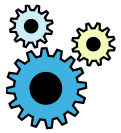





ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Summary of our Approach

1. Knowledge + tractable reasoning is necessary



2. Knowledge acquisition has to be highly automated

3. Large bodies of high-quality knowledge can be acquired from text (“machine reading”) images, and more.

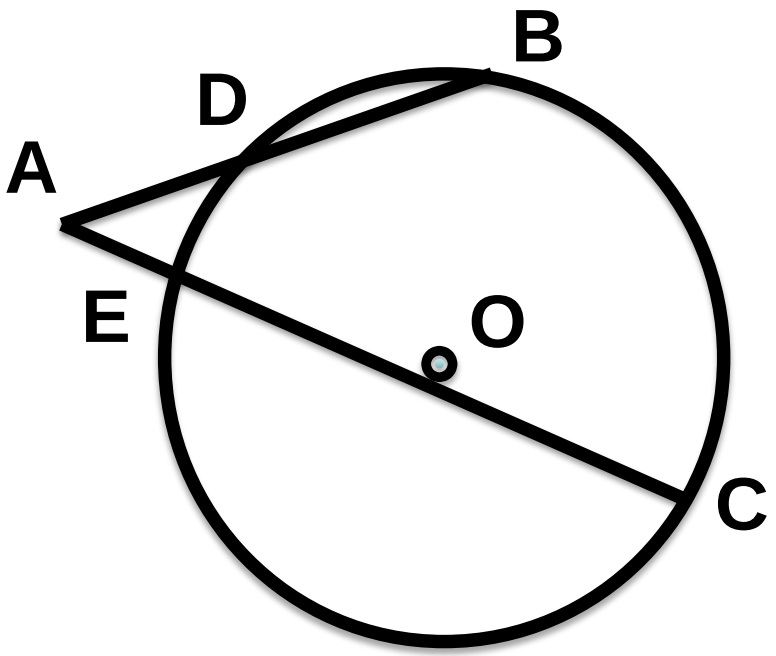
4. Aristo is tested on unseen, standardized, multiple choice tests in science, arithmetic, and geometry

Semantic challenges everywhere…

# Geometry Domain (Case Study)

Combines vision, NLP, and simple semantics
Data set: 100 9th grade geometry problems
**AAAI '14** (**Seo, Hajishirzi, Farhadi, Etzioni**)

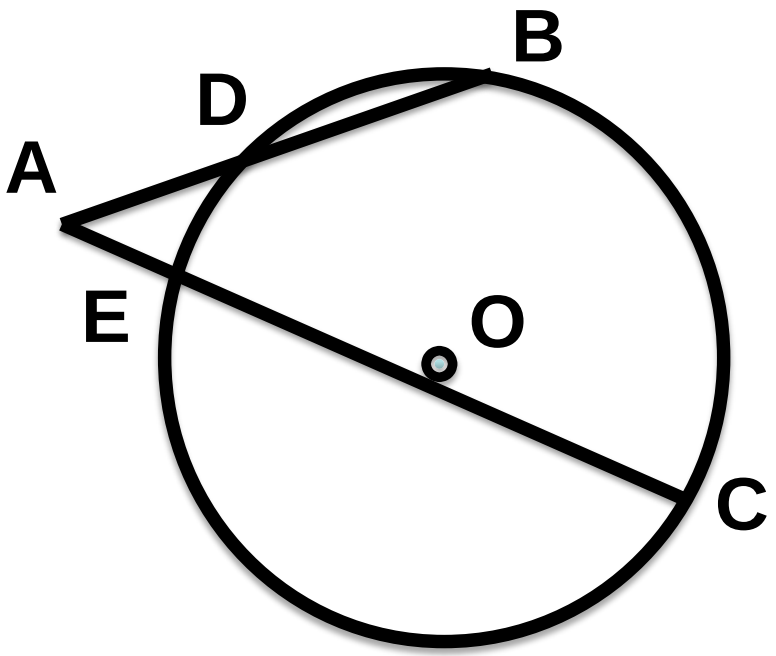
**In the diagram, AB intersects circle O at D, AC intersects circle O at E, AE = 4, AC = 24, and AB = 16. Find AD.**






ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Geometry Diagram Alone






ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Geometry Text Alone

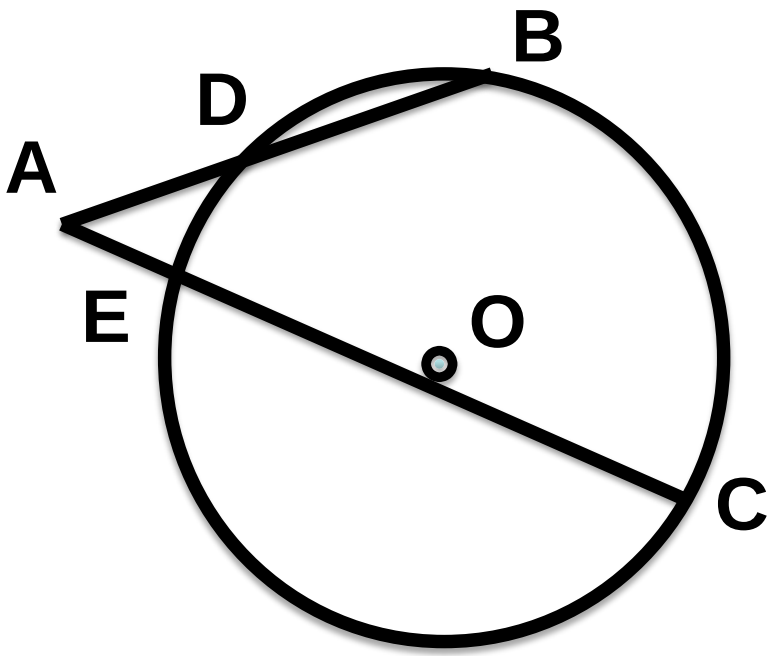
**In the diagram, AB intersects circle O at D, AC intersects circle O at E, AE = 4, AC = 24, and AB = 16. Find AD.**



ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Need Both Diagram and Text

**In the diagram, AB intersects circle O at D, AC intersects circle O at E, AE = 4, AC = 24, and AB = 16. Find AD.**






ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

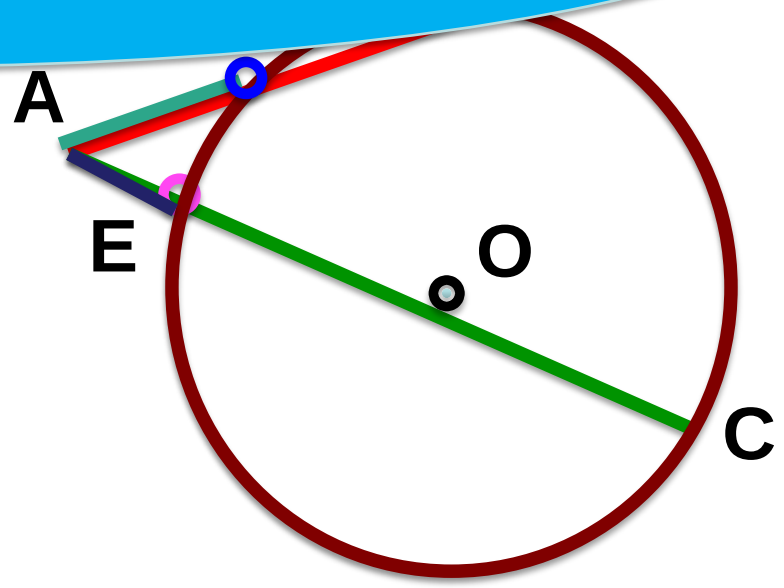
# Diagram Understanding

**Problem:**

1. Identify visual elements
2. Ali

**Diagram understanding ➔ submodular optimization**

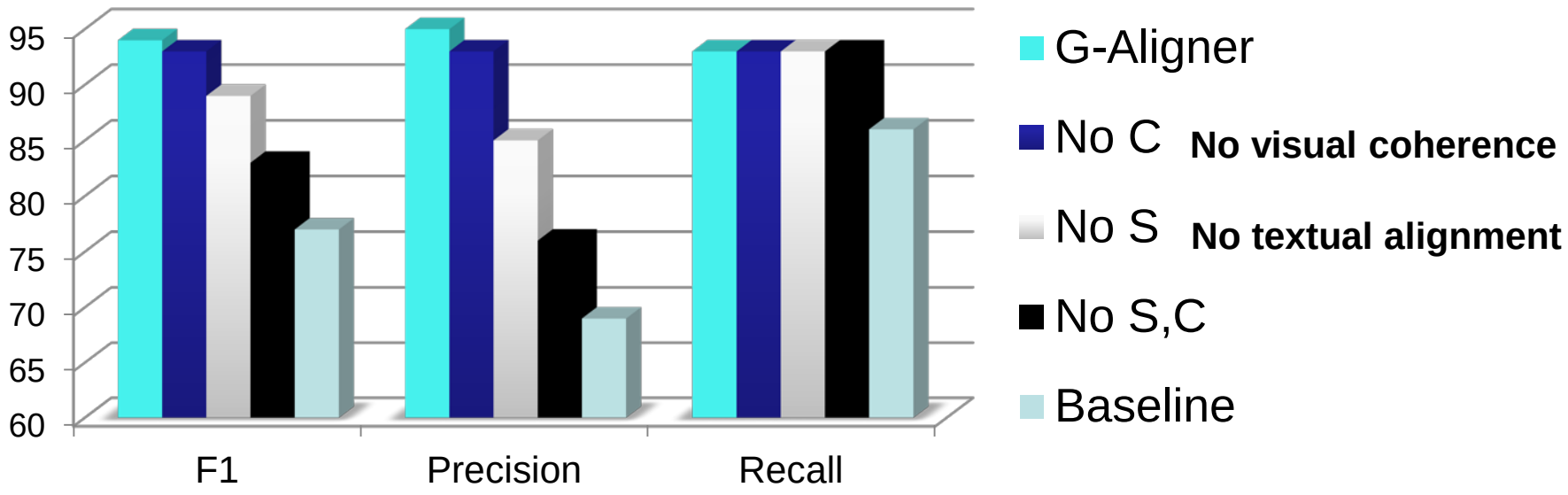
**In the diagra… intersects circle … secant AC intersects circle O at E, AE = 4, AC = 24, and AB = 16. Find AD.**






ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Experimental Study

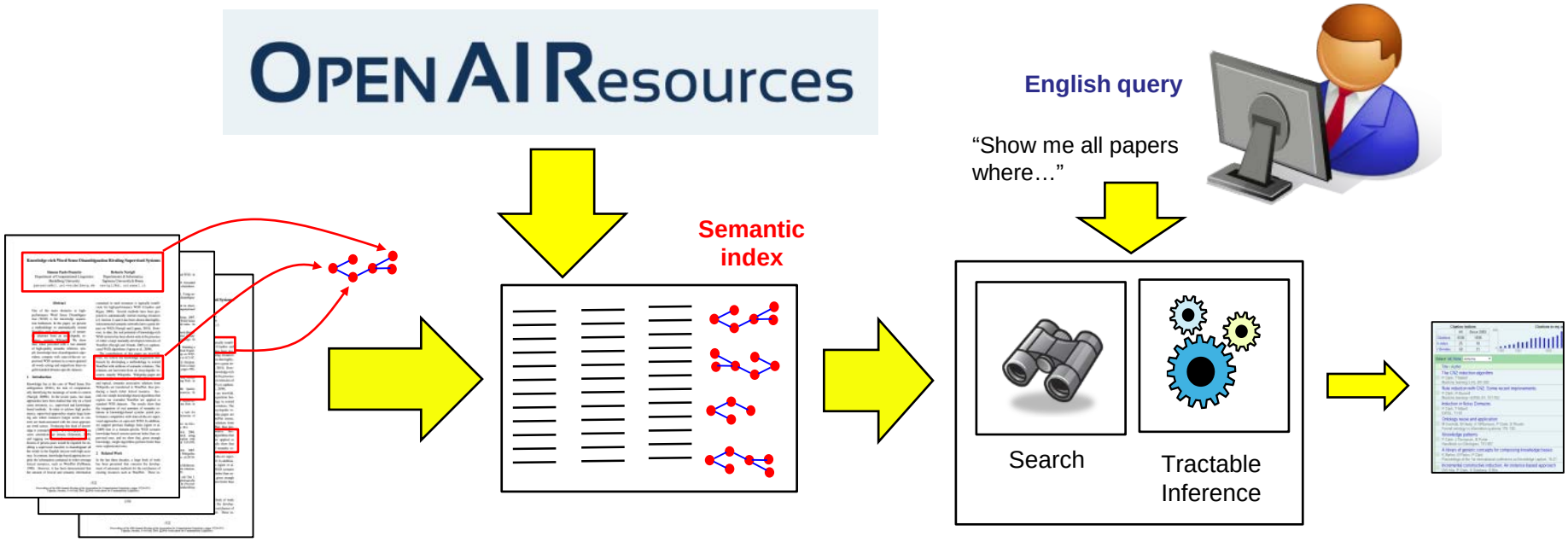





ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# “Semantic Scholar” Project

*"Show all papers where x is implicated in y where the levels of z are elevated"…*




*"lift all scientific boats"*



ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Ambitious Research Investing in Engineering Mindset

**Identify time-saving details early**

**Focus on Tools, viz**

**Design for replicable experiments**



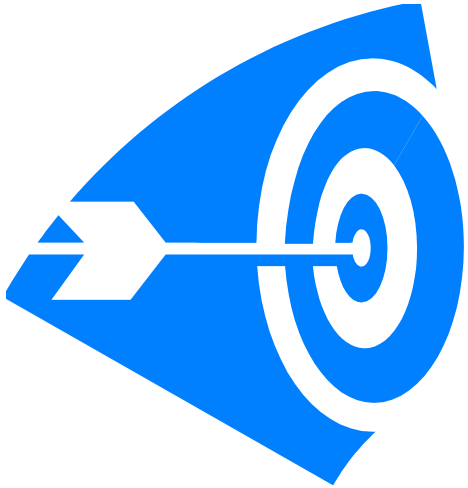
AI2 ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# Summary

- What's Next?  (after "big data" & "deep learning")

***"You can't play 20 questions with Nature and Win!"***

- Focus on data/text Mining systems yielding:
  - Knowledge
  - Reasoning
  - Explanation



# Thank you from AI2!

ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

# What is Data Mining?