Giving Computers Common Sense

Push Singh

MIT Media Lab

Common Sense Computing

9 February 2005

Teaching machines about how people work

- My goal is to build computing systems that have a deep grasp of human nature and the human world.
  - Can understand, explain and predict human physical, social, and mental behavior.
  - Use databases of commonsense knowledge and models of human psychology.
- At the Media Lab we are developing technologies to embed common sense in a broad range of software, devices, and environments — enabling new kinds of intelligent interfaces and applications.

Commonsense-based Applications

A **cell phone** that, although silenced, would know to ring if your mother were to call from the hospital.

A **search engine** that, when you entered “a gift for my baby brother,” would displays a list of children’s toys.

A **personal digital assistant** that would know to cancel a hiking trip with a friend who had broken their leg.

A **camera** that knew, on its own, to take a photo of your sister crossing the finish line at a marathon.

Commonsense-based Applications

- People who are badly hurt may go to the hospital.
- We want those we care about to be healthy.

- A gift should be something the recipient would like.
- Children like to play with toys.

- Hiking can be a relaxing activity.
- People with broken legs have difficulty walking.

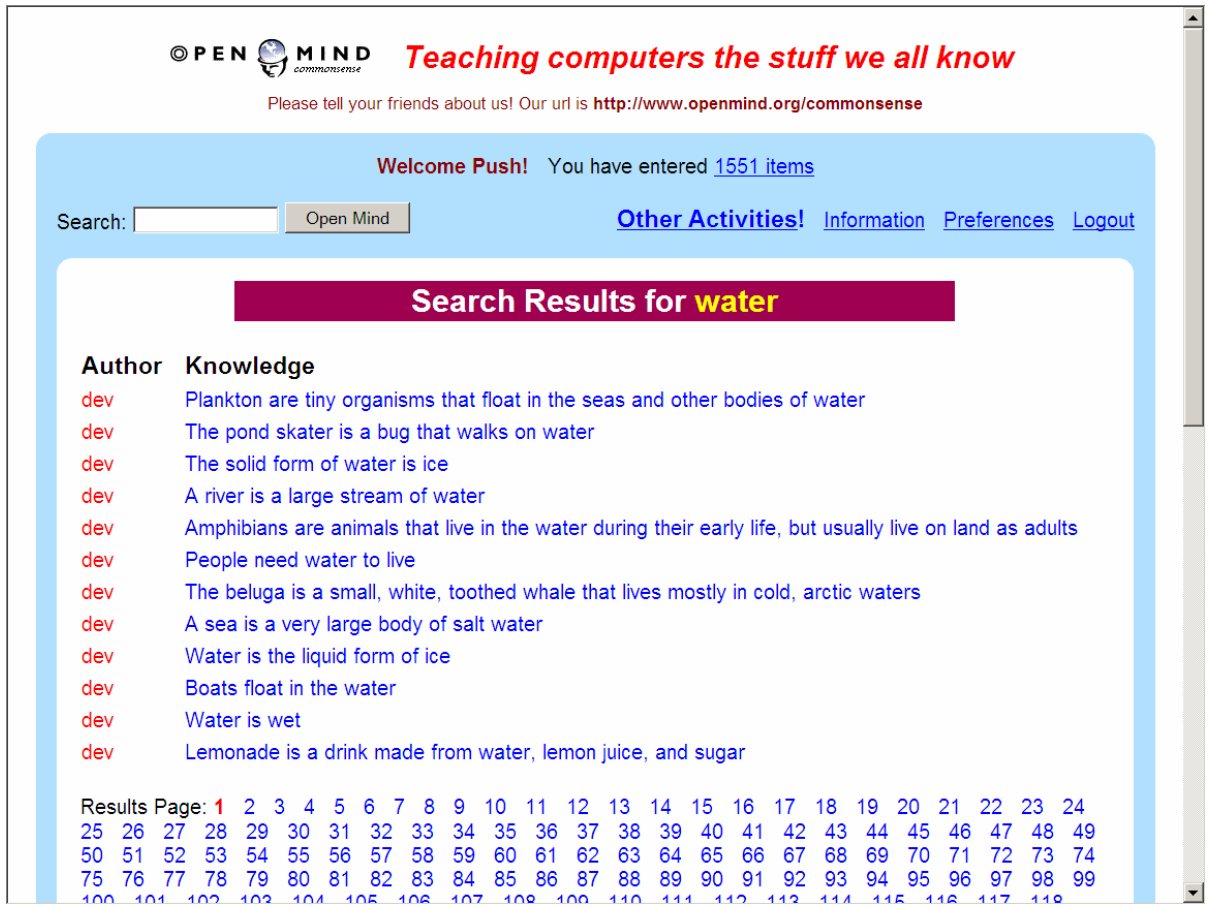
- Cameras are for recording events people find significant.
- Crossing the finish line is an significant event.

Can we take a Wikipedia approach to teaching machines common sense?

- We need to build commonsense databases with tens of millions of elements. How can we do this?
- Observation: *Everyone has the common sense we want to give our machines.*
- Can we build a system that learns from tens of thousands of ordinary people?

Open Mind: Common Sense

750,000 facts from 15,000 people



Push Singh (2002). **The public acquisition of commonsense knowledge.** *Proceedings of AAAI Spring Symposium on Acquiring (and Using) Linguistic (and World) Knowledge for Information Access.* Palo Alto, CA.

Simple acquisition activities

Please enter a general fact (not a caption) this picture reminds you of:

tools can be used to fix things

Teach Open Mind!

What is the relationship between the word **eat** and the word **cook**?

people often cook food before eating it

Teach Open Mind!

Please connect the following two sentences with as ***few intermediate sentences*** as possible:

Birds in the wild often fly.

Birds can sing

A songbird can wake you in the morning

You eat breakfast soon after you wake up

The first thing you do when you eat breakfast is make coffee

Teach Open Mind! | Try different sentences

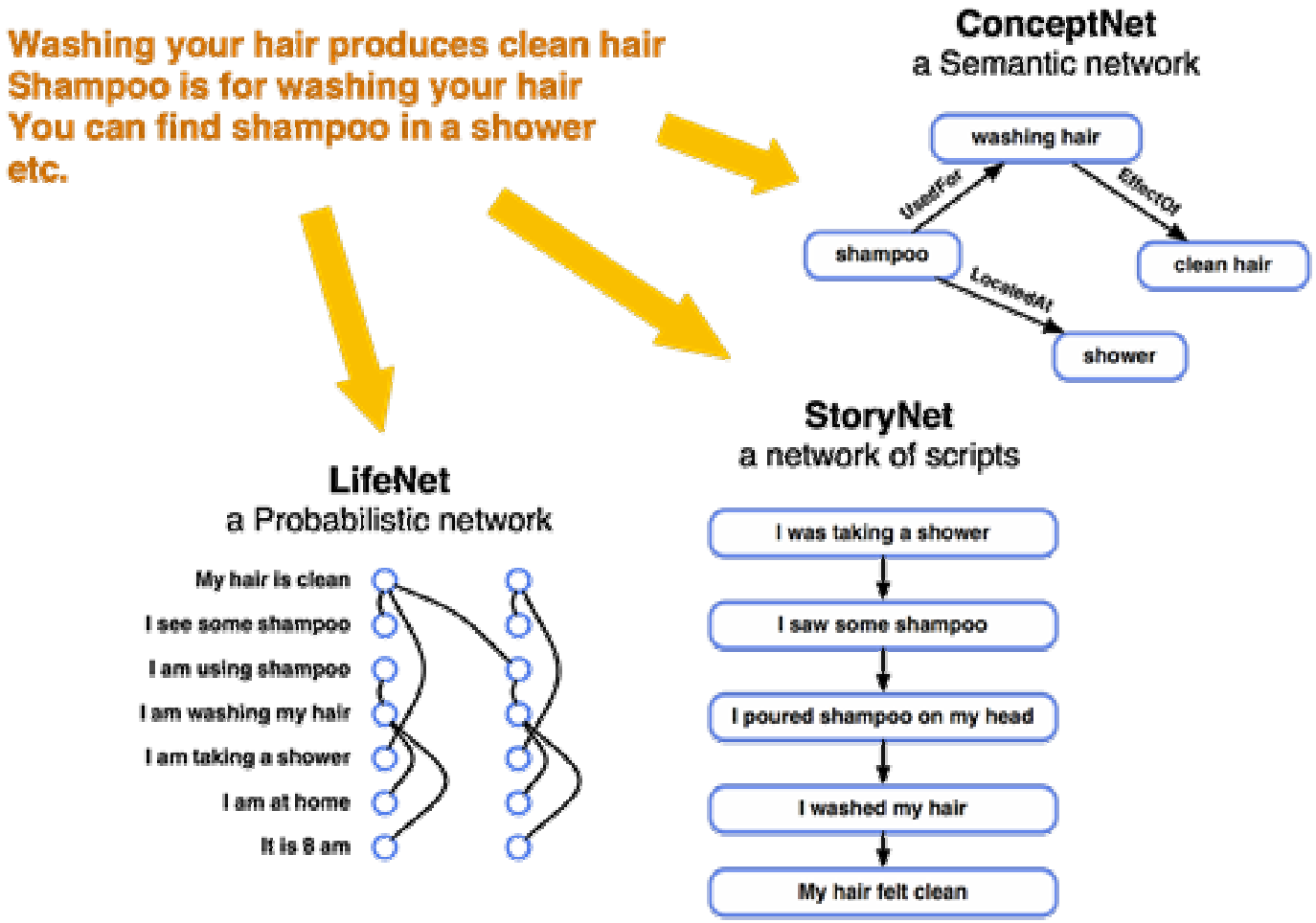
Something that might happen as a consequence of going jogging is you get sweaty

Teach Open Mind!

It must be full of a bunch of crank submissions, vandalism, and plain old sophomoric stupidity. But it's not. It's not half bad. In places, and increasingly, it's of very high quality. And that's even more paradoxical.

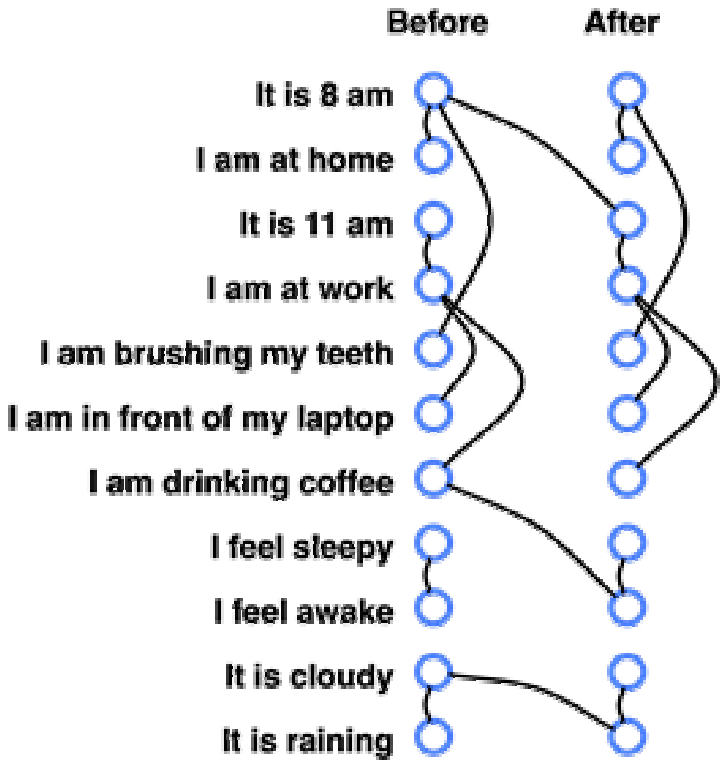
– Larry Sanger, Wikipedia Co-founder

Extracting to multiple representations



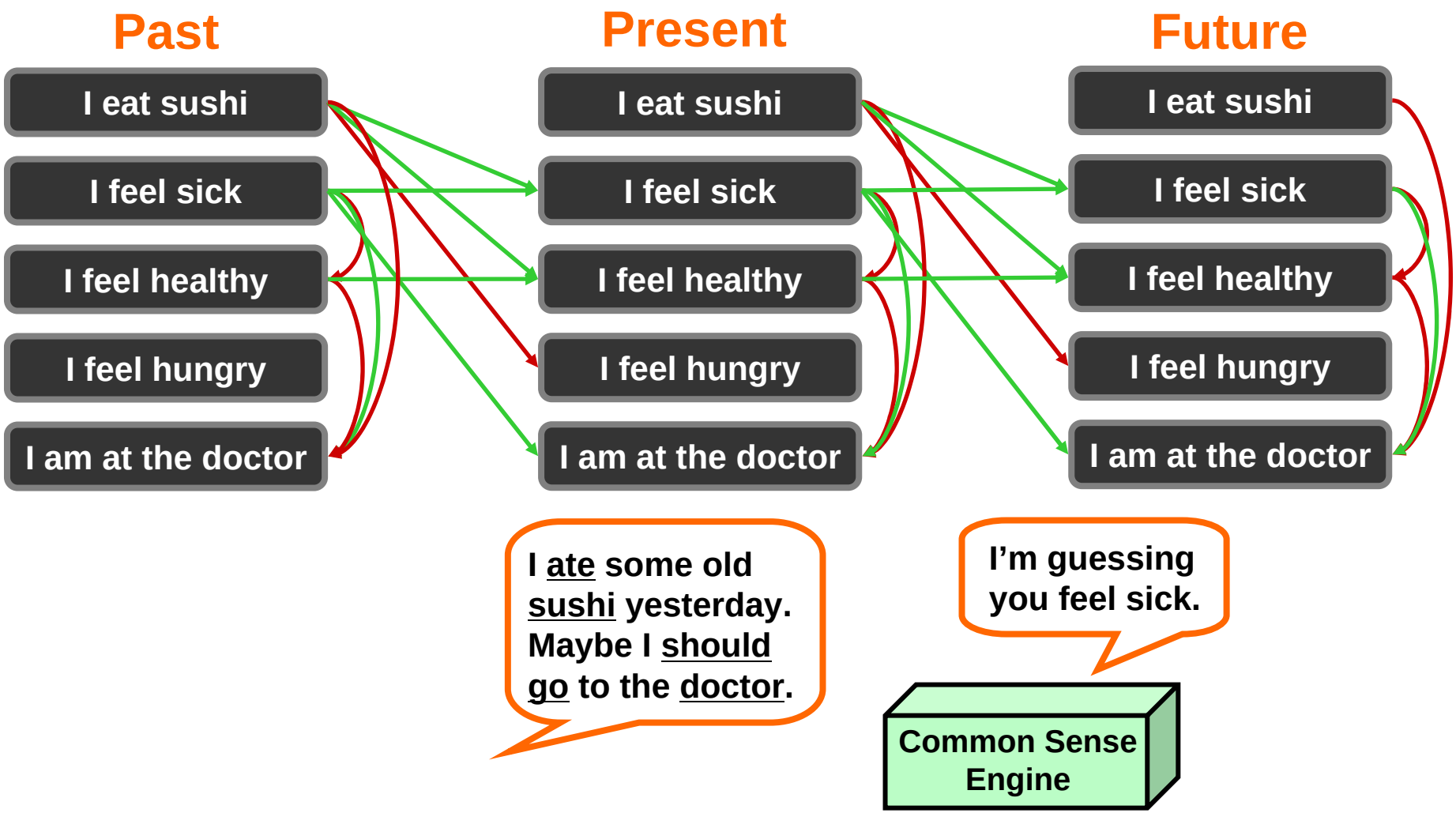
Push Singh, Barbara Barry, and Hugo Liu (2004). **Teaching machines about everyday life.** *BT Technology Journal*, 22(4):227-240.

LifeNet:
a 1st-person probabilistic model of human experience
400,000 links relating 100,000 propositions



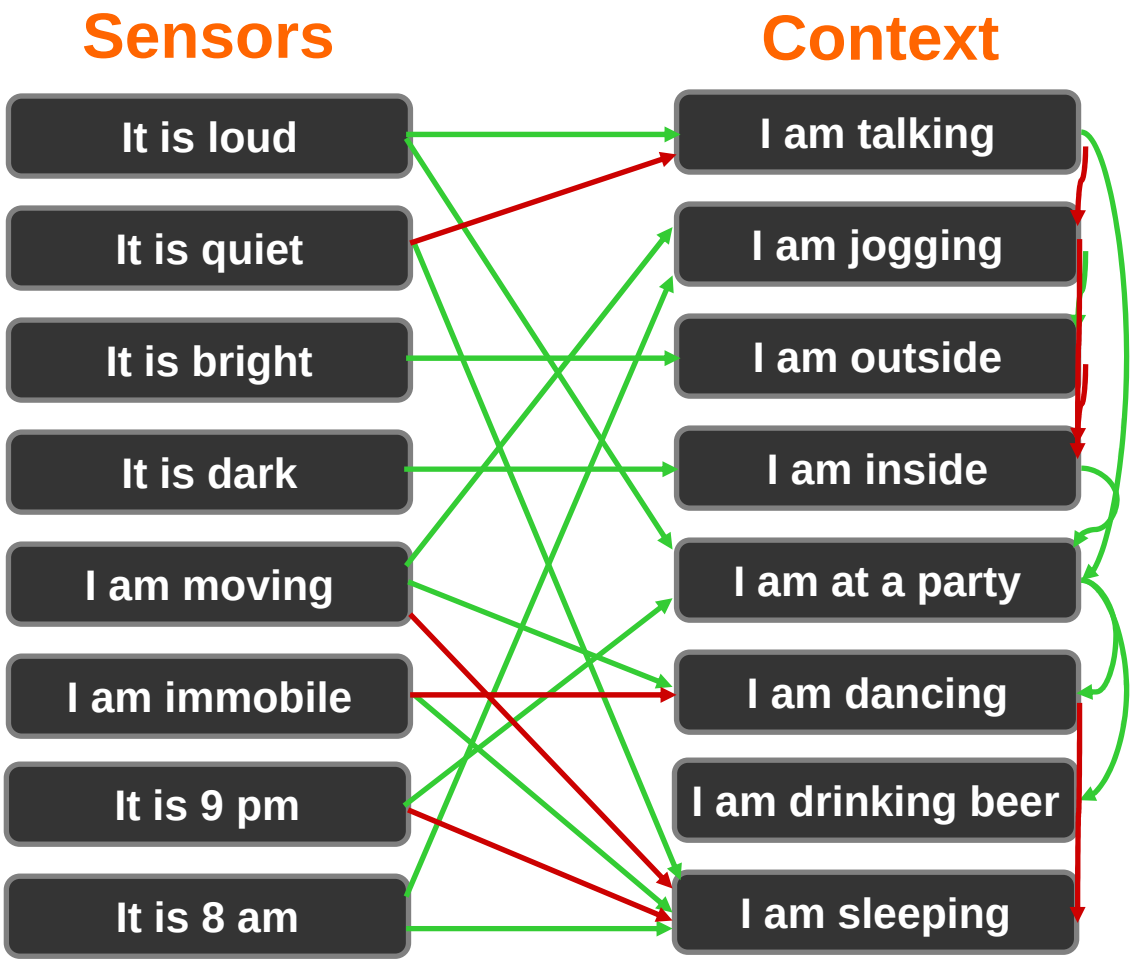
Push Singh and William Williams (2003). **LifeNet: a propositional model of ordinary human activity.** *Proceedings of the Workshop on Distributed and Collaborative Knowledge Capture (DC-KCAP) at K-CAP 2003*. Sanibel Island, FL.

Example: Inferring context from speech



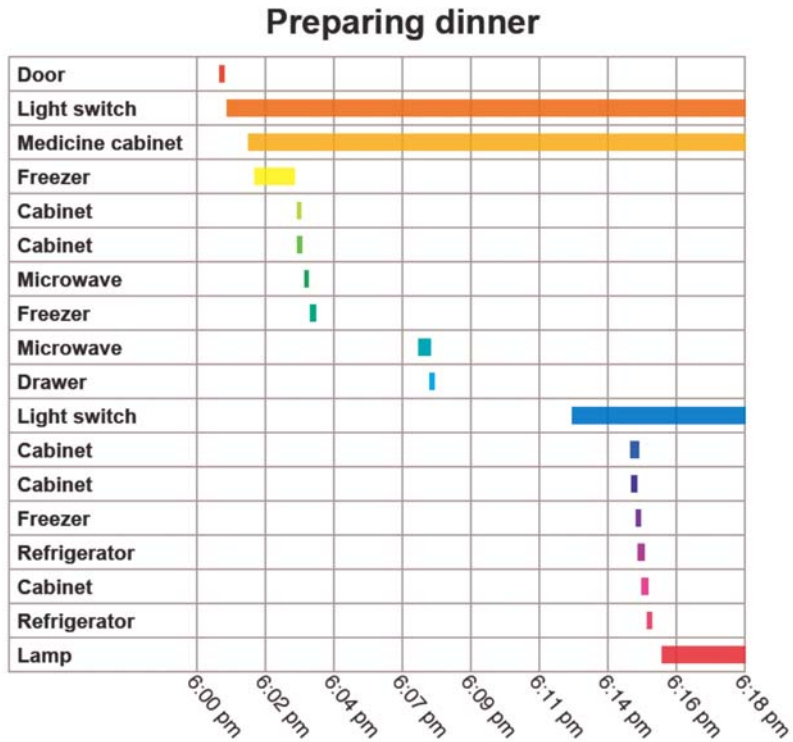
Nathan Eagle and Push Singh (2004). **Context sensing using speech and common sense.** *Proceedings of the NAACL/HLT 2004 workshop on Higher-Level Linguistic and Other Knowledge for Automatic Speech Processing.*

Example: Inferring context from sensors



Learning from and interpreting streams of sensor data

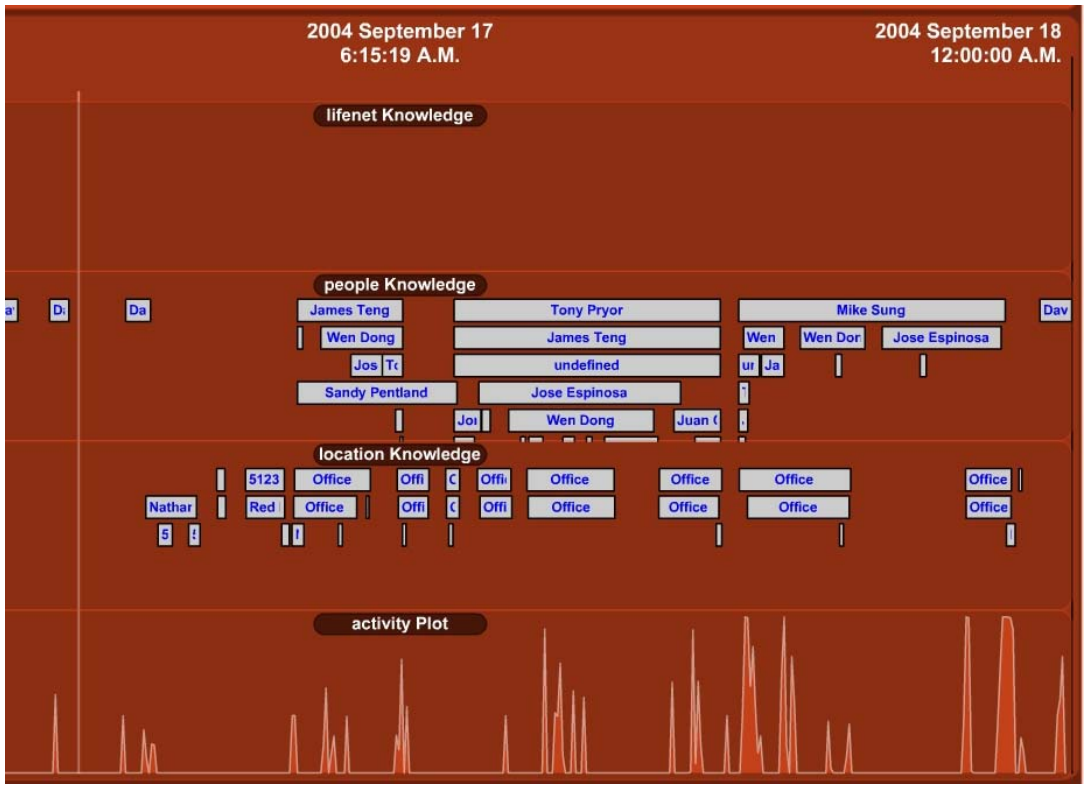
PlaceLab:
a sensor-rich apartment



E. Munguia Tapia. **"Activity Recognition in the Home Setting Using Simple and Ubiquitous Sensors"**. S.M. Thesis, Massachusetts Institute of Technology, 2003.

Learning from people as they live their lives

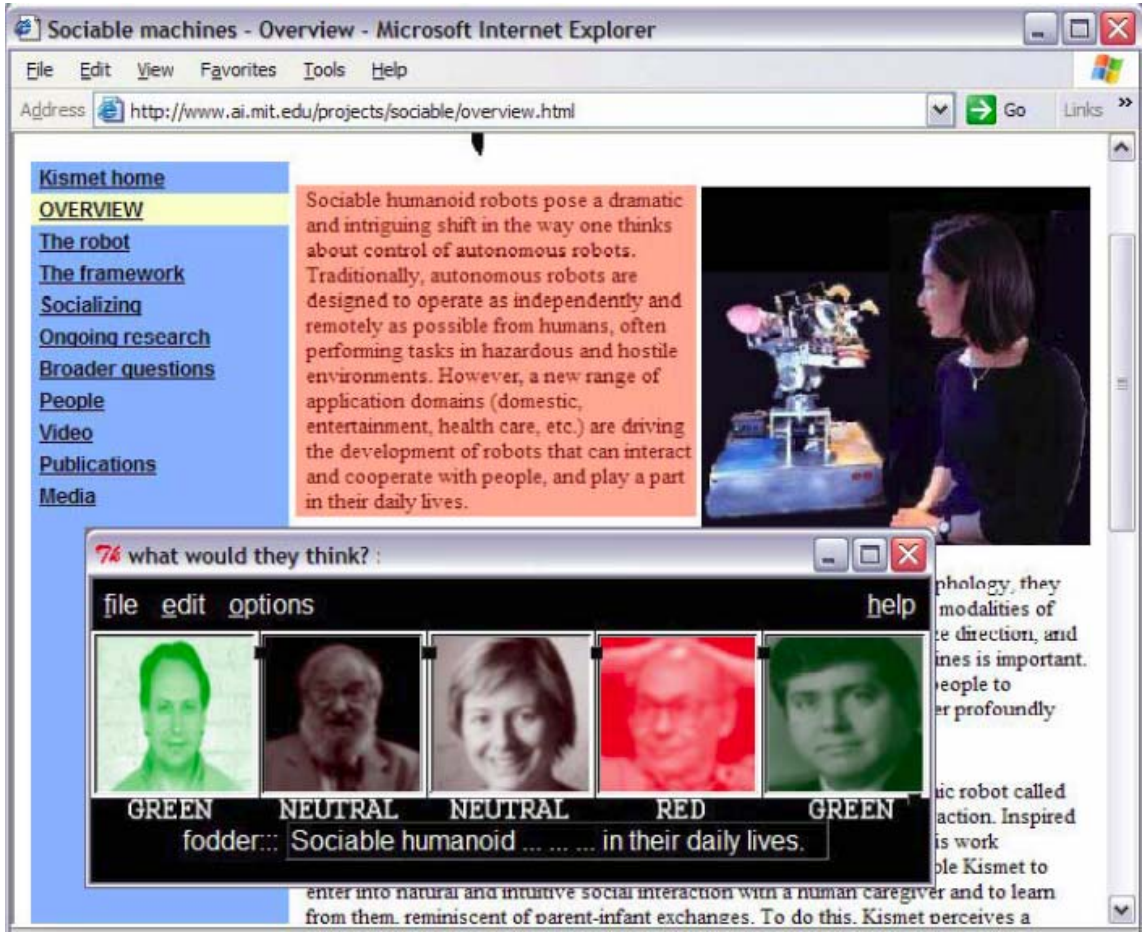
100 cell phone users around MIT



what was I doing?
who was I with?
what was I feeling?

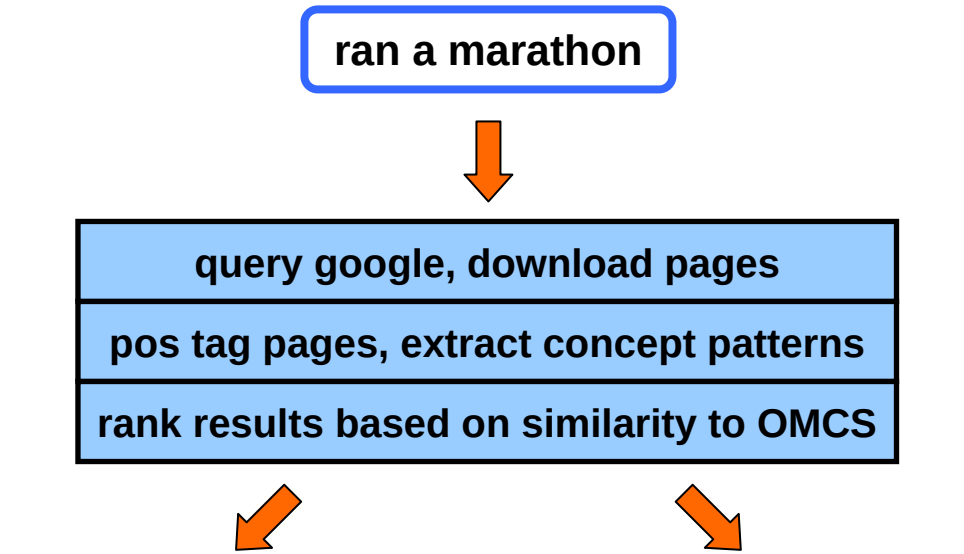
what will I be doing next week?
who might I run into?
what should I do to prepare?

What would they think?



Hugo Liu and Pattie Maes (2004). **What Would They Think? A Computational Model of Attitudes**. *Proceedings of the ACM International Conference on Intelligent User Interfaces, IUI 2004*, January 13–16, 2004, Madeira, Funchal, Portugal. ACM 2004, ISBN 1-58113-815-6, pp. 38-45.

ConceptMiner:
Can we find common sense on the web?



good results (30%)
is an important event
running event
finish a marathon
complete a marathon
run for pleasure
could barely run two miles

poor results (70%)
has run a total
are optional run days
marathon the adventure
Not only does marathon
marathon a reality
say running this marathon

Mining Results

Query	Mined	Clean
ran a marathon	1614	556
watched a marathon	1354	94
finished a marathon	297	94
trained for a marathon	265	94
entered a marathon	259	83
won a marathon	157	42
Totals	4221	1033

1 million good results
per machine-day

Our projects

- LifeNet (temporal probabilistic model)
- ConceptNet (large-scale semantic net)
- StoryNet (structured story knowledge base)
- GoalNet (typical human goals and priorities)
- SituationNet (prototypical situations)
- ShapeNet (shape kb for visual commonsense)
- GlueNet (connecting representations)
- ThinkNet (reflective reasoning with stories)
- ComicKit (telling stories by writing online comics)
- Serendipity (learning behavior from experience)
- ConceptMiner (terascale web mining)
- EM-ONE (implementing the Emotion Machine)

Representing knowledge in multiple ways

StoryNet

LifeNet

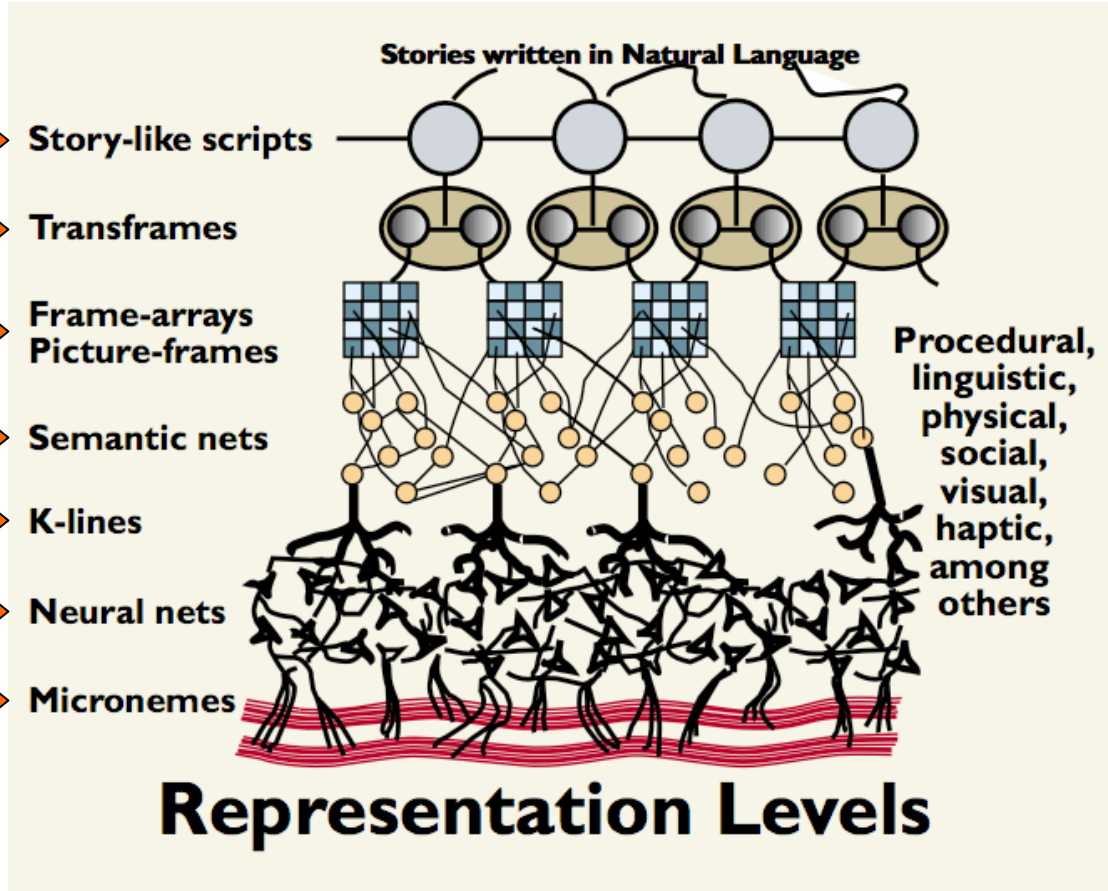
ShapeNet

ConceptNet

SituationNet

?

?



An architecture for the mind

how to think

Self-Conscious Thinking
Concerned with relationship between this mind and others, including self-appraisal by comparing one's abilities and goals with those of others.

Self-Reflective Thinking
Concerned with larger scale models of "self", including the extent and boundaries of one's physical and cognitive abilities and knowledge.

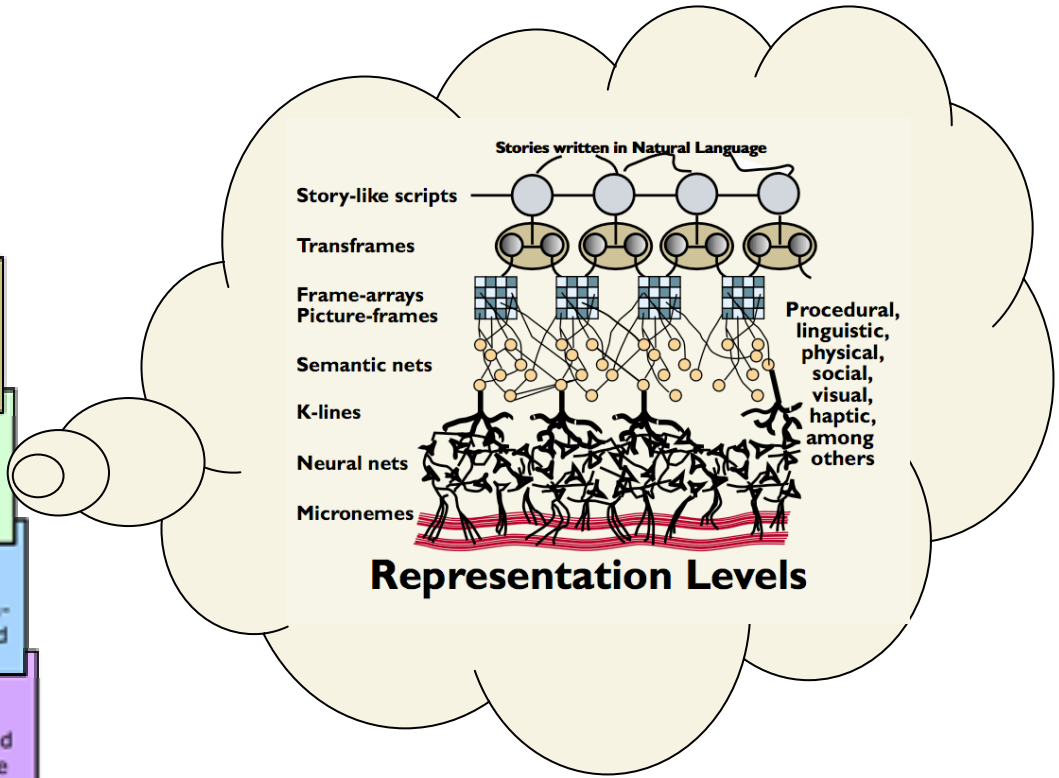
Reflective Thinking
Reflects on and manages deliberative activity, including assigning credit to inference methods, selecting suitable representations, and so forth.

Deliberative Thinking
Reasons about the situations and events in the external world, e.g. prediction, explanation, planning, diagnosis, generalization.

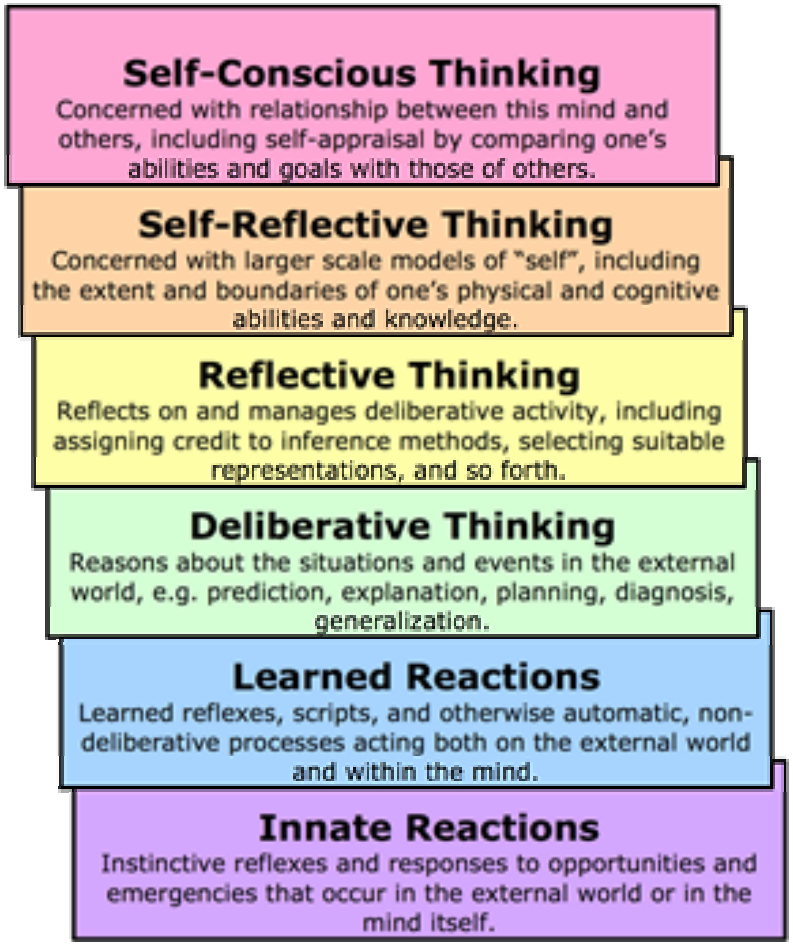
Learned Reactions
Learned reflexes, scripts, and otherwise automatic, non-deliberative processes acting both on the external world and within the mind.

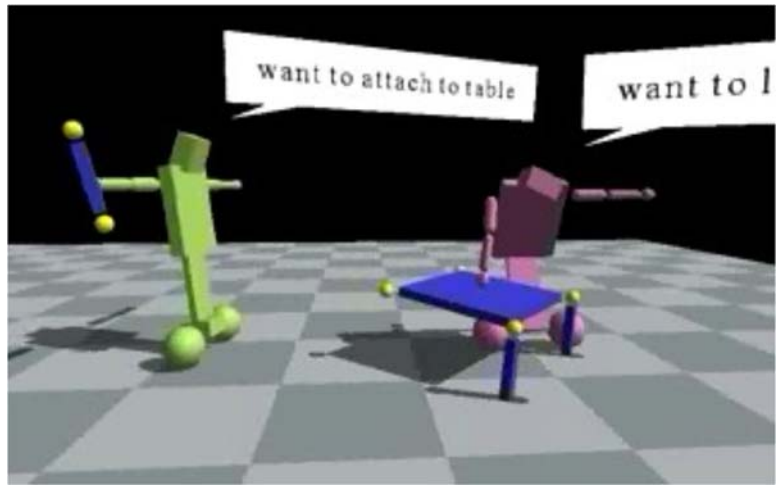
Innate Reactions
Instinctive reflexes and responses to opportunities and emergencies that occur in the external world or in the mind itself.

what to think



A reflective and self-aware architecture for intelligence





Questions the architecture can ask itself:

What is the best thing for me to do now?
What will happen next following this event?
What would explain why this event occurred?
What can I learn from this failure?
What might go wrong while performing this action?
What could be it the negative consequences?
Why are they taking that action?
etc.

Push Singh and Marvin Minsky (2003). **An architecture for combining ways to think.** *Proceedings of the International Conference on Knowledge Intensive Multi-Agent Systems.* Cambridge, MA.

Conclusions

1. Using the web, we have been able to give computers a first cut at a broad-spectrum understanding of the human world.
2. We are searching for ways to use “messy” knowledge bases, e.g. using probabilistic representations.
3. We are connecting commonsense reasoning directly to the real world through sensory interfaces, to learn from and also to help interpret that sensory data.
4. We are developing cognitive architectures that are **self-reflective** and **multi-representational**, to try to achieve something closer to human-like thinking.
5. We are developing many applications based on these technologies. See *Beating Common Sense into Interactive Applications* (AI Magazine, Winter 2004).

More information

http://csc.media.mit.edu

Contributors:

Barbara Barry, Walter Bender, Tim Chklovski, Nathan Eagle, Ian Eslick, Jose Espinosa, Ashwani Kumar, Henry Lieberman, Hugo Liu, Erik Mueller, Marvin Minsky, Bo Morgan, Alex Pentland, Push Singh, and Ryan Williams

how much does a person know?

- Can you estimate the number of…
  - **words** you know and things you know about those words
  - **activities** you are familiar with and their structures
  - **bodily postures** you can get into and what you can do in them
  - **object appearances** you can recognize from different angles
  - **layouts of typical places** you know and what you can do there
  - **actions** you can predict the effects of in different contexts
  - **functions** various shaped objects serve
  - **goals** you want to achieve–both large and small
  - **emotional states** you enter and how they affect your thinking

- It seems likely that the number is in at least the millions and possibly there are tens or even **hundreds of millions** of items of common sense.

The Challenge: Ambiguity

“the sky is blue”

You can always make things more precise. Does this mean…

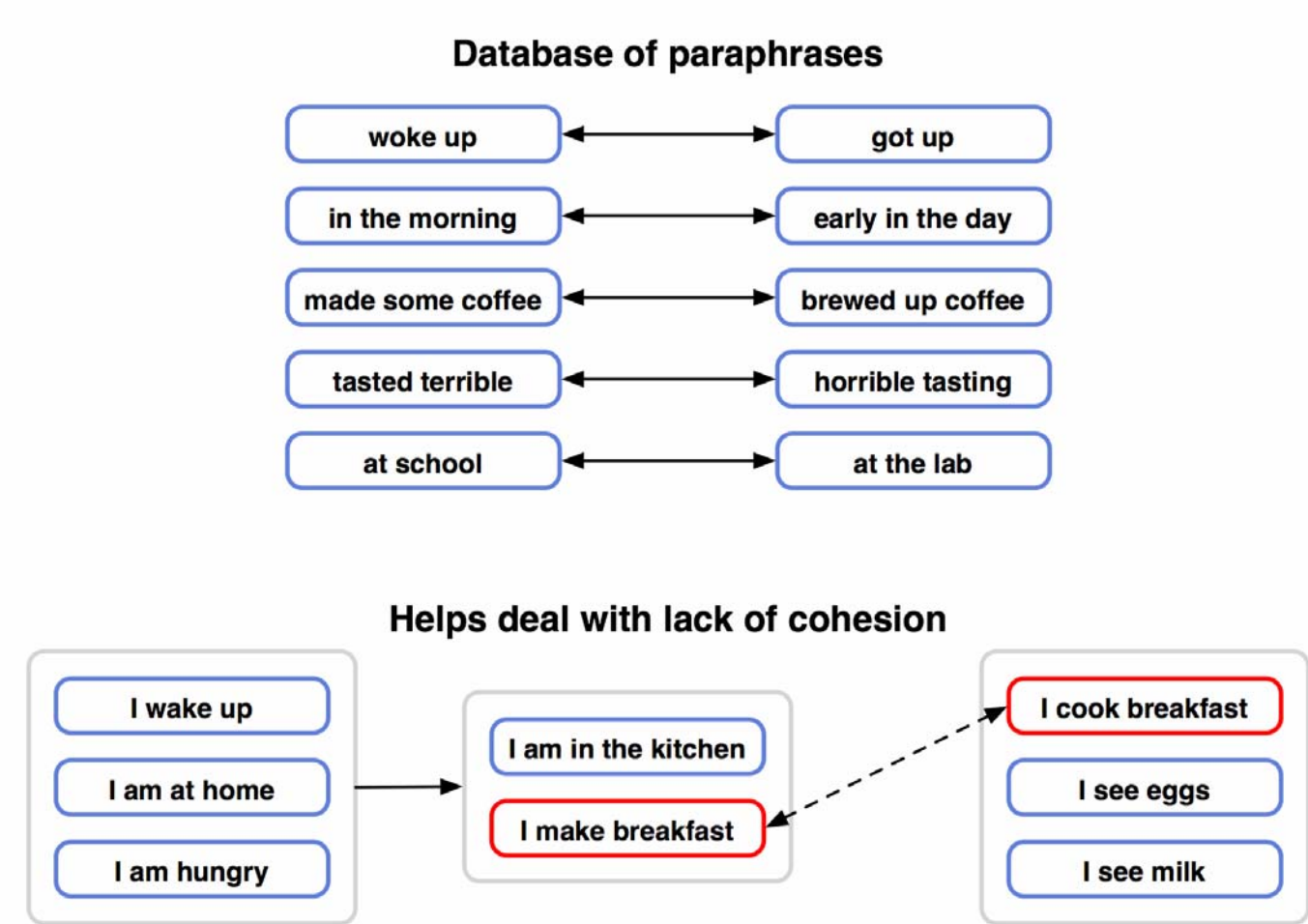
… the sky is always blue?
… the sky is blue only sometimes (i.e. during the day)?
… the sky is mostly blue but has other colors?
… the sky is always blue but the things in it have other colors?

How can we reason with this kind of ambiguity?

do we need a more precise ontology under the hood?
can we learn the distribution of more precise meanings?
can contexts help disambiguate, as is done in NLP?
maybe we can leave representations ambiguous?
(ambiguous senses sometimes help by lending additional meaning)

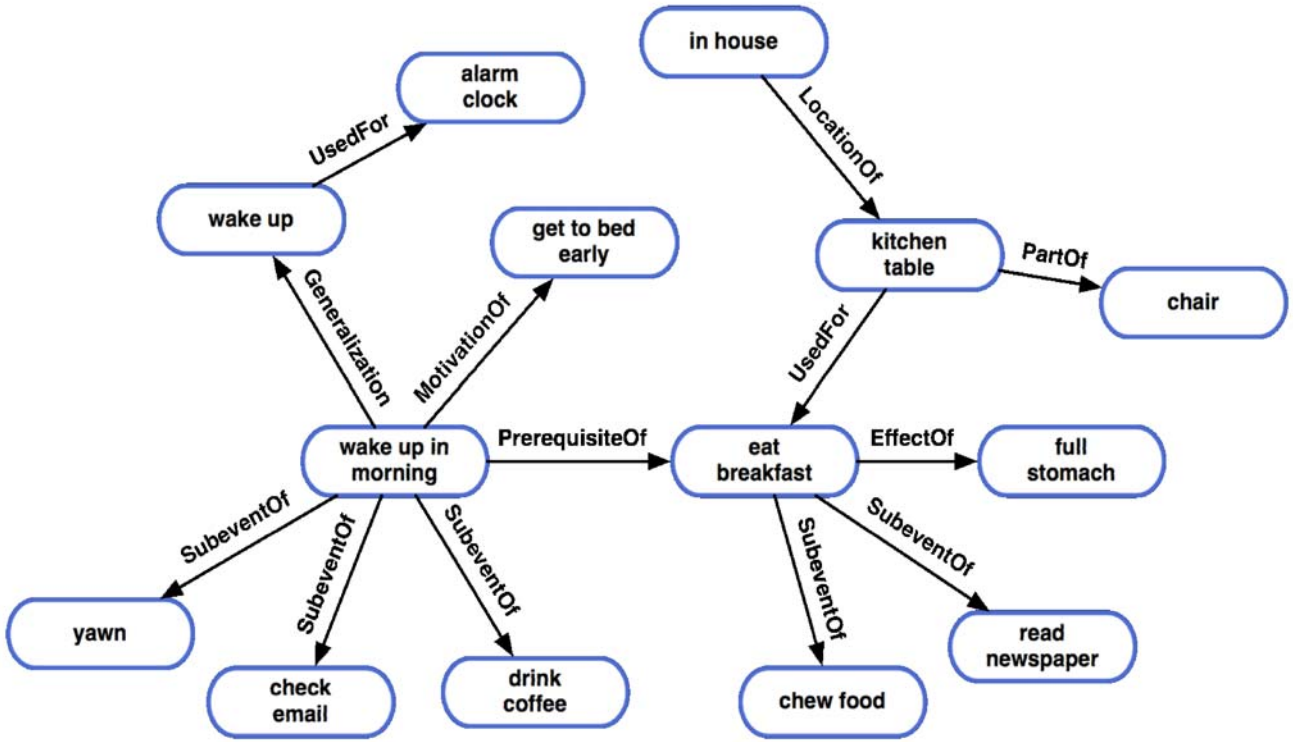
Ambiguous representation involve fewer symbols and less syntactic complexity, and are simpler to think about. This makes building KBs much easier, at the cost of their being harder to apply.

GlueNet:
Tools for Ontology Alignment



ConceptNet: a large semantic network

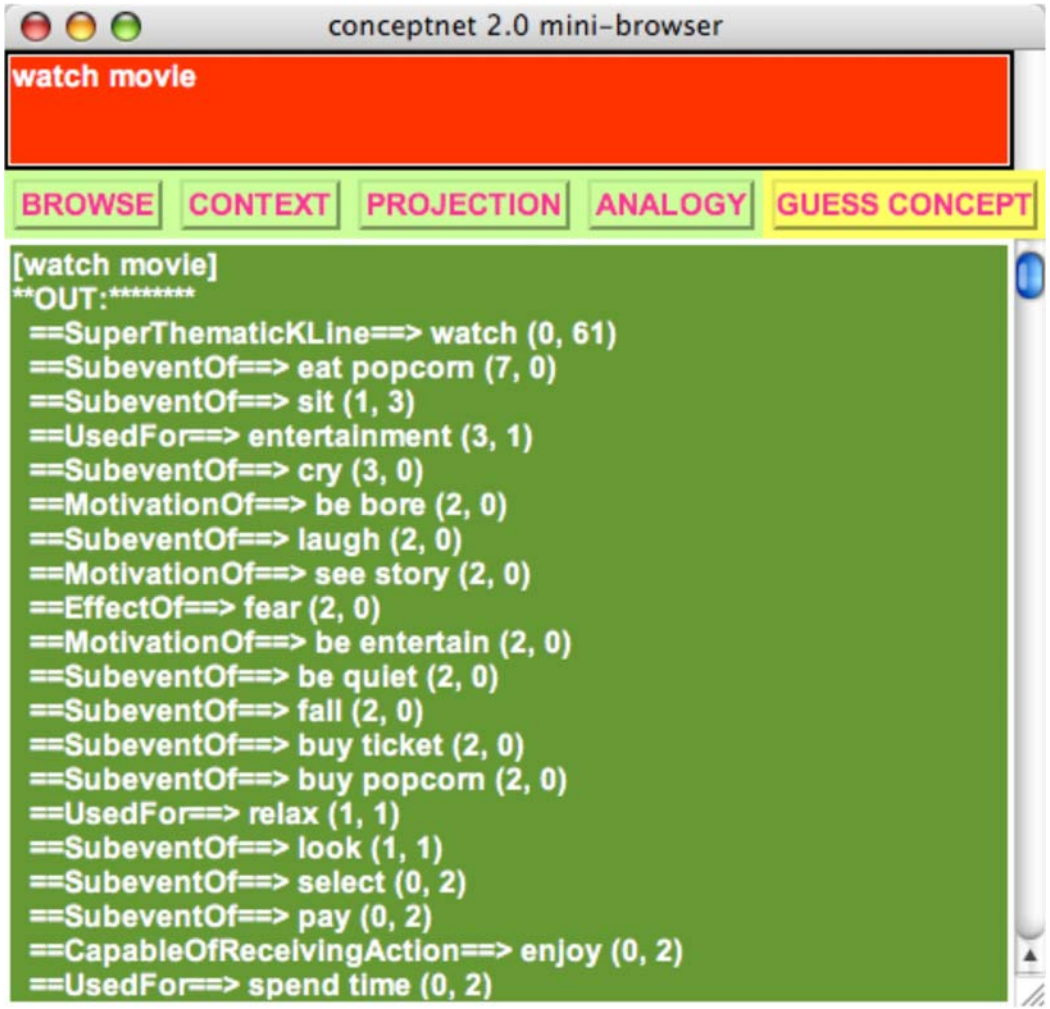
1.6 million links relating 300,000 concepts



Hugo Liu and Push Singh (2004). **ConceptNet: a practical commonsense reasoning toolkit.** *BT Technology Journal*, 22(4):211-226.

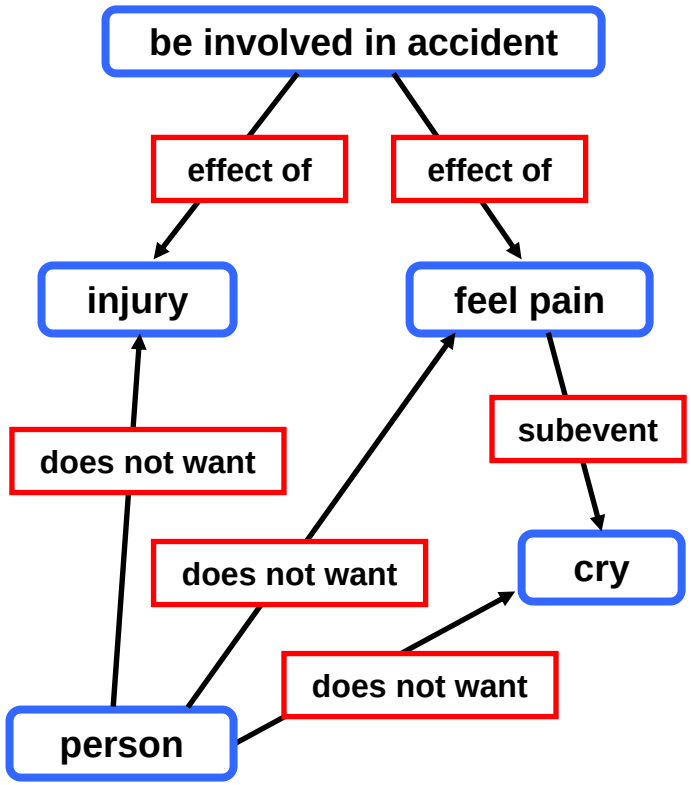
ConceptNet toolkit

(www.conceptnet.org)



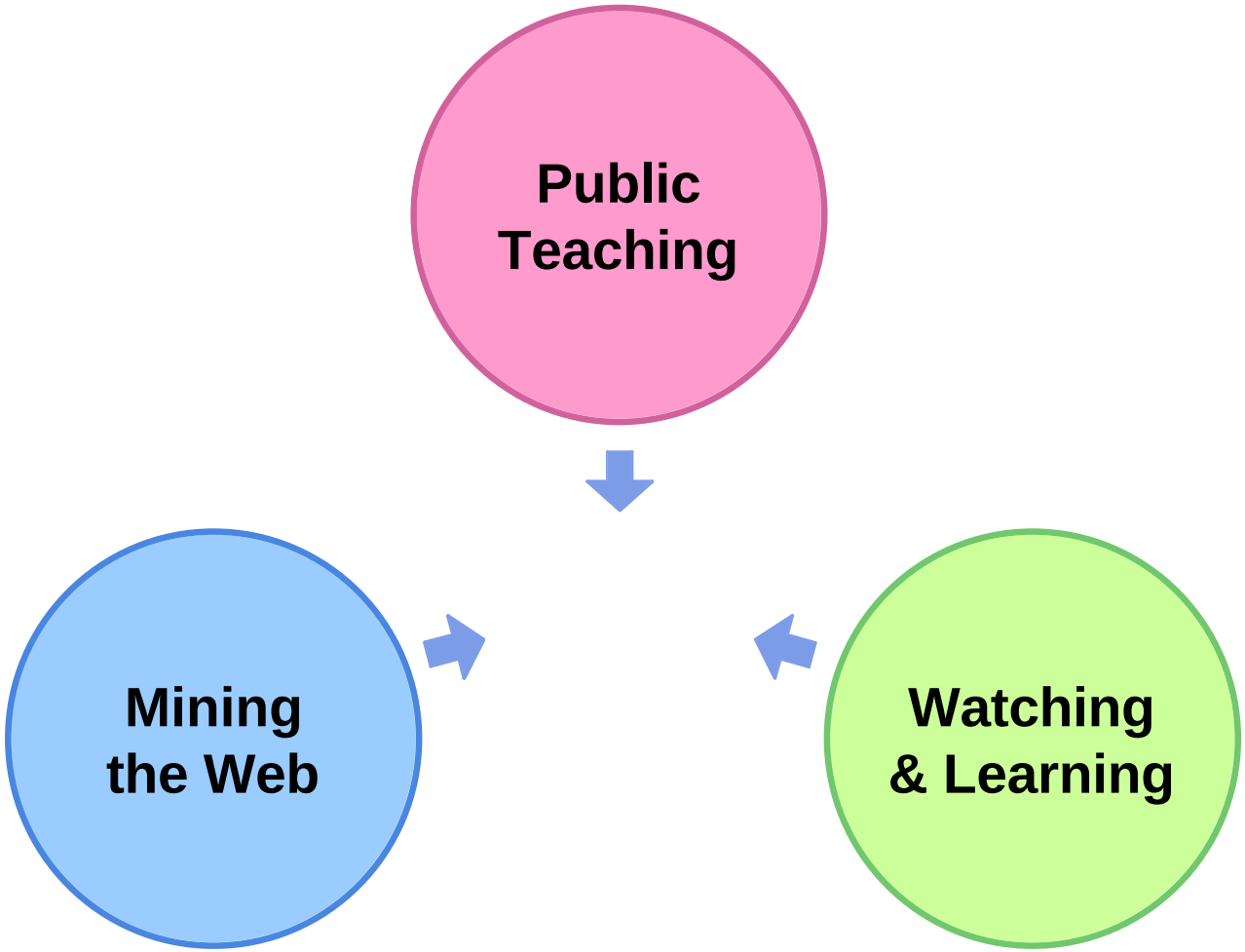
- runs as XML-RPC server
- integrated NLP system (part-of-speech tagging, chunking)
- functionality
  - topic-jisting (e.g. an article with concepts, gun, convenience store, demand money and make getaway might suggest the topics “robbery” and “crime”)
  - affect-sensing (e.g. this email is sad and angry)
  - text summarization
  - and more
- versions in Python, Java, Common Lisp, C, Ruby

example: inferring affect of text

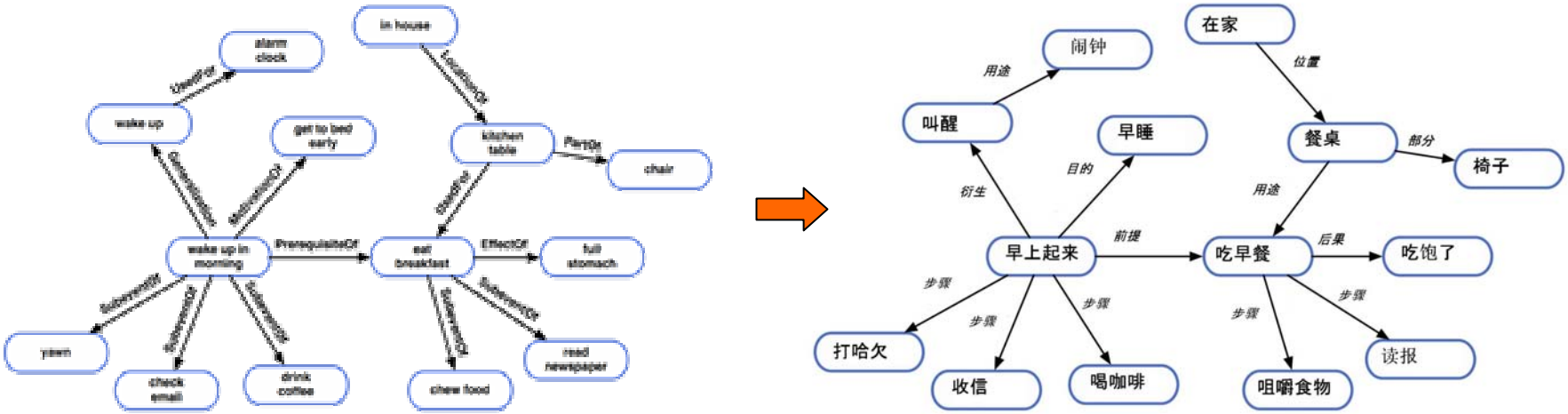


Hugo Liu, Henry Lieberman, and Ted Selker (2003). **A Model of Textual Affect Sensing using Real-World Knowledge.** *Proceedings of the Seventh International Conference on Intelligent User Interfaces (IUI 2003)*, pp. 125-132. Miami, Florida. **Outstanding Paper Award**.

Multiple Ways to Acquire Knowledge



Multilingual ConceptNets



95% of ConceptNet is valid in Chinese culture.
Developing a tool to make it easy to translate nodes.
Also: Spanish, Japanese, Hindi

SituationNet:
detailed descriptions of situations

Example:
Gisting fine-grained topics from speech

Actual Topic

What to get for lunch in the cafeteria (streaming data to an access point mapped as ‘cafeteria’.)

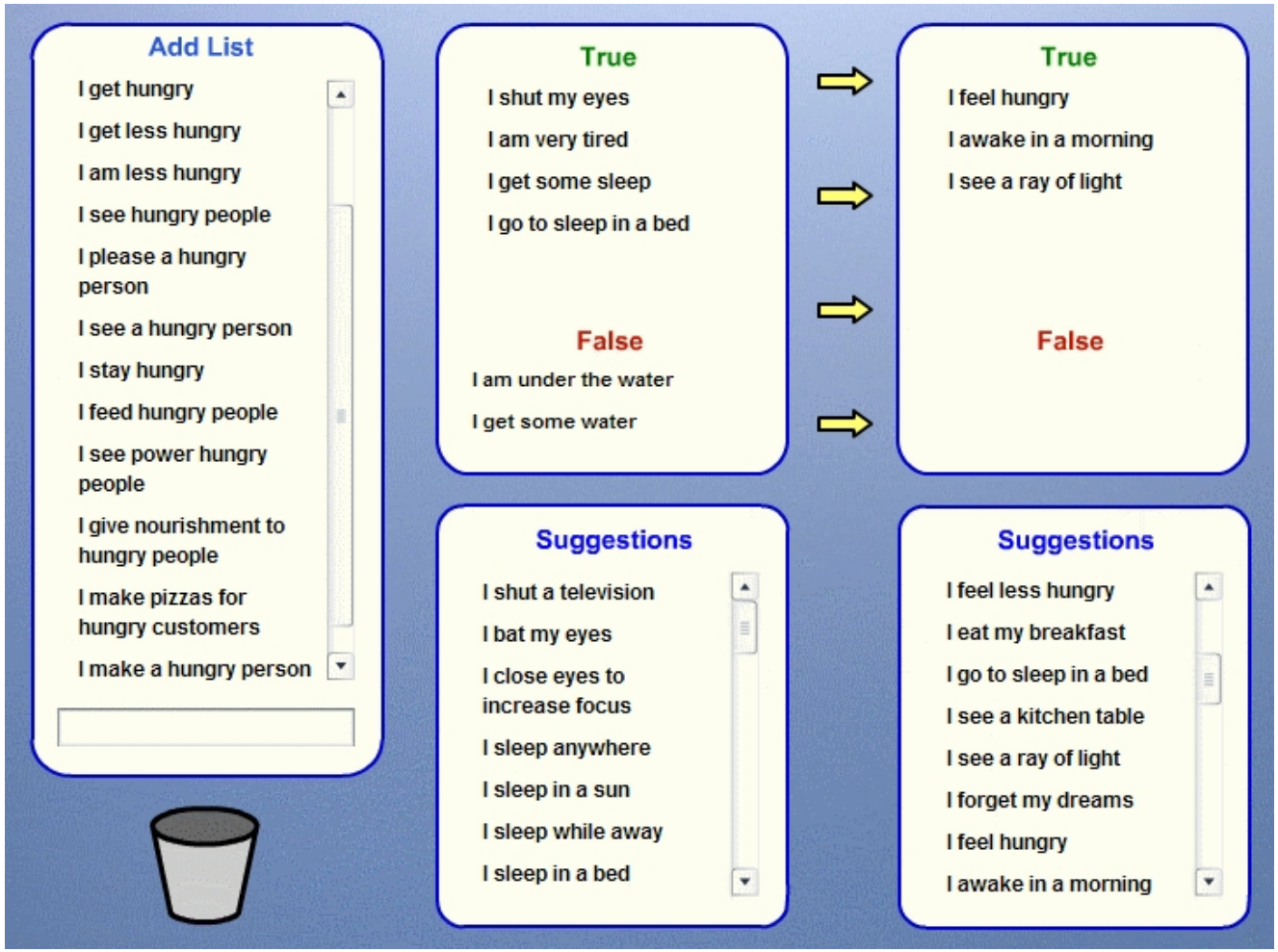
Transcription

Store going to stop and listen to type of its cellular and **fries** he backed a **bill** in the one everyone get a guess but that some of the past like a **salad bar** and some offense militias cambers the site **fast food** them and the **styrofoam** large **chicken nuggets** son is a pretty pleased even guess I as long as can’t you don’t have to wait too long its complicity sunrise against NAFTA pact if for **lunch**

Without Context	With Context
(5) talk with someone far away	(27) eat in fast food restaurant
(5) buy beer	(21) eat in restaurant
(5) eat in restaurant	(18) wait on table
(5) buy hamburger	(16) wait table
(4) go to hairdresser	(16) go to restaurant
(4) wait in line	(15) know how much you owe restaurant

Nathan Eagle, Push Singh, and Alex (Sandy) Pentland (2003). **Common sense conversations: understanding casual conversation using a common sense database.** *Proceedings of the Artificial Intelligence, Information Access, and Mobile Computing Workshop (IJCAI 2003).* Acapulco, Mexico.

Adding to LifeNet

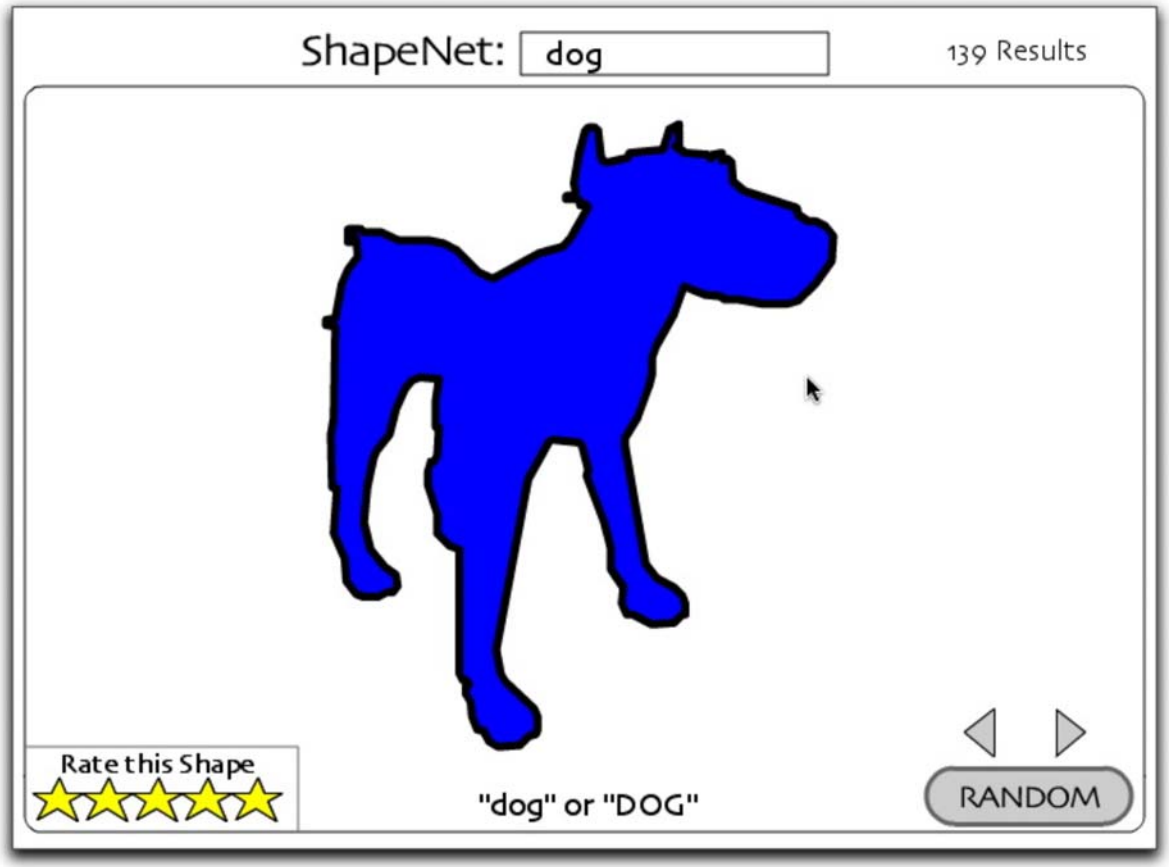


Commonsense Computing @ MIT Media Lab

We are developing a suite of tools:

- *Commonsense knowledge bases*
  - By enlisting ten thousand people over the web
  - By mining millions of pages on the web
  - By watching people as they live their lives
- *Commonsense reasoning systems*
  - That are tolerant to ambiguity and errors
  - Based on multiple representation schemes
- *Commonsense computing architecture*
  - Marvin Minsky's new Emotion Machine architecture

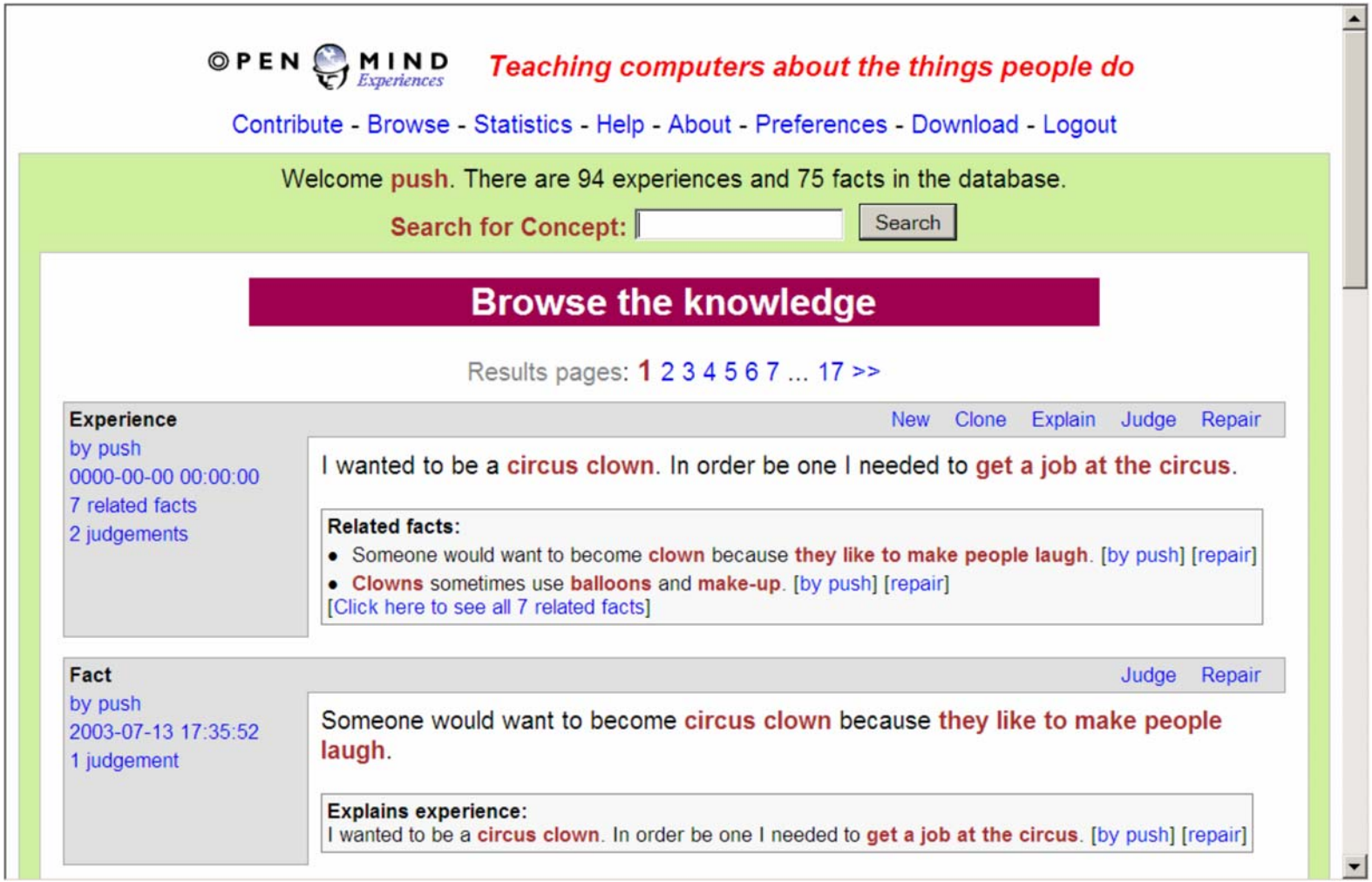
ShapeNet: Spatial Common Sense



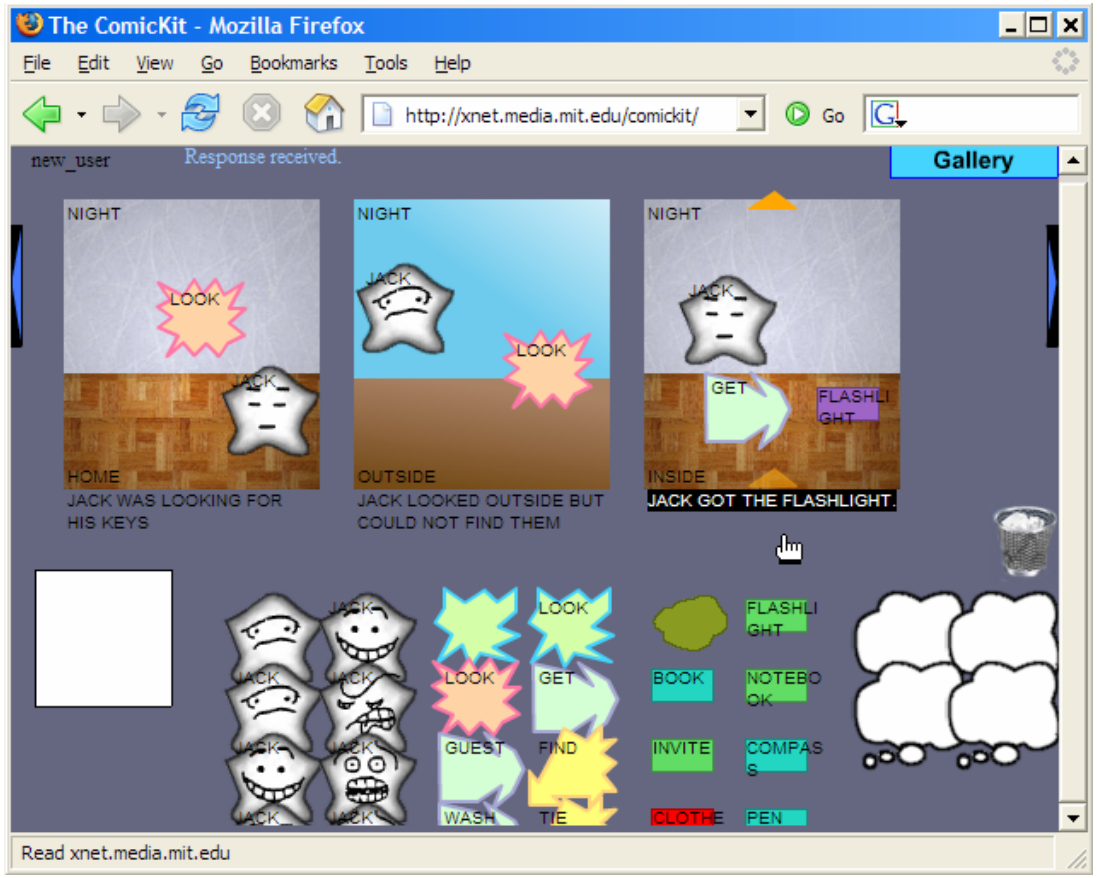
StoryNet


I felt bored
I decided to go the beach
I got in my car
I drove to the beach
It was very crowded
I swam in the surf
I got tired

Collecting Stories



Acquiring Commonsensical Stories from Kids



Ryan Williams, Barbara Barry, and Push Singh (2005). **ComicKit: acquiring story scripts using commonsense feedback.** *Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI 2005).* San Diego, CA.

The Structure of a Commonsense Appliance

Understand the situation
- **Watch what you do**
- **Listen to you speak**
- **Monitor sensors**
- **Read your blog**

Take useful actions
- **Acts on your behalf**
- **Suggests things to do**
- **Warns you of problems**
- **Remind you of things**

All systems will come with embedded common sense!

What are you doing and why?
What can you do to achieve your goals?
Is there a better situation to be in this one?
What just happened and what might happen?
What might go wrong?
Could someone else help?

interpreting sensor data using common sense

- Given a partial and raw sensory stream, we can try to interpret it at a higher level.
- One possible narrative explanation:
  - person enters room
  - person gathers ingredients
  - person begins cooking
  - person prepares ingredients
  - person cooks those
  - person gets knife
  - etc.
- Using a maximum likelihood framework -- given observations, infer most likely high events.