Knowledge Representation

CS 395/495 - Section 24

Ken Forbus

Fall 2003

Outline

- Contact information
- What this course is about
- Motivation
- Syllabus
- Projects
- Tools
- Your first assignment

Contact Information

- The course web site has almost everything:
  - **Lecture notes, assignments, announcements, supplementary materials, …**
  - **http://www.cs.northwestern.edu/~forbus/cs395-kr/index.html**
- Staff cs395-kr-staff@cs.northwestern.edu
  - Instructor: Ken Forbus
  - TA: Praveen Paritosh
- On-line discussions will use NU's CMS system https://courses.northwestern.edu/webapps/login

What this course is about

- You will learn how to represent knowledge formally
  - So precisely that computer programs can use it
- You will learn state of the art representation schemes for core kinds of knowledge
  - Space, time, quantity, events, causality…
- You will learn how to build representations in the context of large, pre-existing knowledge bases
  - There's no need to start from scratch anymore
  - Like living on top of other people's software, there are skills to learn to do this well

Why learn about knowledge representation?

- A brief history of knowledge representation
- The Revolution is now!
  - For developing broad-scale cognitive models
  - For developing radically new kinds of AI systems

Prehistory

- Philosophers started it all
- Attempts to formalize argumentation
  - Rhetoric
- Thought as calculation
  - Boole's Laws of Thought
  - Leibnitz and others
- Sowa's book on KR has good historical material

Early AI enthusiasms

- Logic and theorem proving eagerly adopted
- Computational issues forced consideration of how to package up knowledge, control inference
  - Frame languages
  - Special-purpose KR languages
- More noise and heat than light in early debates
  - Formalists versus Hackers

Form minus content

- Movement in 1980s: KR = Formal KR
  - Reaction to lack of clear semantics
  - Identification of formality with precision
  - Focus on general logical schemes, not specific domains
- Consequences
  - Lots of technical progress
  - Common perception of sterility in many areas, e.g. non-monotonic logics
  - Most exciting KR work didn't appear in KR community, e.g., qualitative physics, CYC project, ...

The Dark Years

- Various anti-representation/representation lite trends declare the end of KR
  - Connectionism, Behaviorism, Fuzzy logic, Feature-based representations
  - All fine things as part of a balanced approach, but look out for those boom and bust cycles…
- As limitations become clearer, expectations become more realistic
- Fads turn into serious research efforts, willing to interact with other approaches

Why one needs serious KR

Q: How much folic acid should an expectant mother take daily?

A: 360 tons

Q: What is the diameter of the Earth?

A: 14 ft.

Q: How many states have a lottery?

A: 3,312

Examples drawn from TREC9 systems, which use word-level information retrieval and statistical information and heuristics to guess answers to questions

The Representation Resurgence

- Representation Lite hits too many walls
  - NL frontier now integrating semantics with statistical techniques
- Dramatic success stories in narrow areas
  - Scheduling: Desert Shield, I2, Detecting money laundering via link analysis, …
- Steady scientific progress in AI
  - KR now embracing content again
- Moore's law is making it all practical
  - A few hundred MB for a large KB versus 1GB for one hour of video seems like a bargain.

The Revolution is happening now

- Ideas, technologies, and tools now coming together
- Clear perception arising of need for common sense knowledge bases
  - Keeping up with the Web -- NLP rises again!
    - See Semantic Web, DAML projects
  - Software that you treat as a collaborator
  - Knowledge management
- The infrastructure is being created today
- Those who understand KR will shape what happens

New territory for Cognitive Science

- Most cognitive theories make assumptions about representation
  - Sometimes explicit, sometimes implicit, often untested
- Most cognitive simulations tested only with small descriptions, generated expressly for that purpose
  - Fine for early testing, but more robust testing would be useful
- Large, off-the-shelf knowledge bases will make new kinds of cognitive simulations possible

Example: Large-scale lexical databases

- Miller *et al*'s WordNet
  - Heavily used by AI community
  - http://www.cogsci.princeton.edu/~wn/
- Fillmore *et al*'s FrameNet
  - Deeper analysis of verbs
  - http://www.icsi.berkeley.edu/~framenet/
- Palmer's VerbNet
  - Cross-linguistic from the start
  - Used in NL interactions with virtual humans
  - http://www.cis.upenn.edu/~mpalmer/project_pages/VerbNet.htm

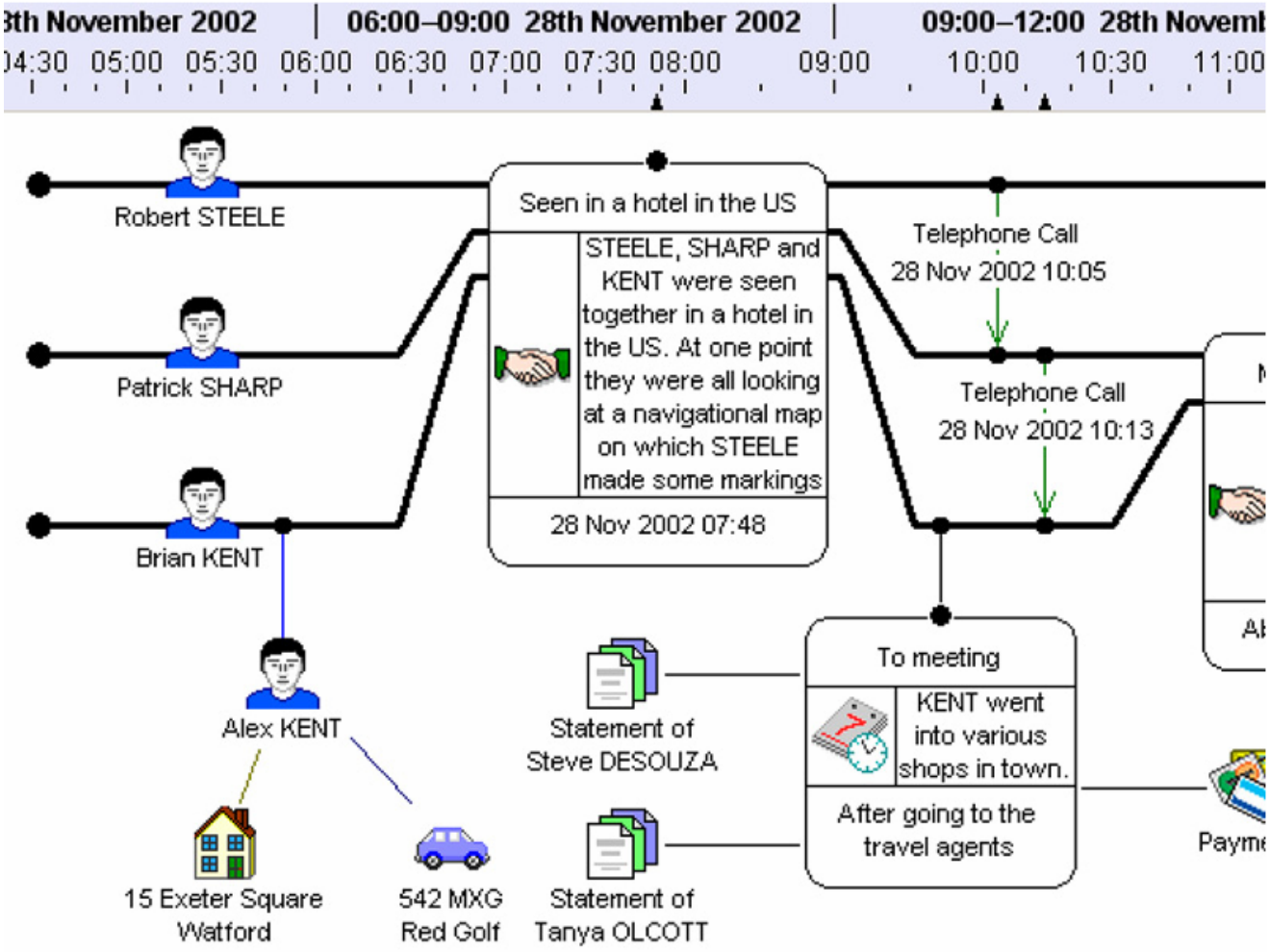
Example: Natural Language Understanding

- Sven Kuehne's NLU system built by combining and extending existing systems
  - Allen's parser
  - Comlex lexicon
  - Cyc KB NL semantics & concepts
  - Framenet conventions
  - Qualitative Process theory

"The pressure in the cylinder increases." ➔

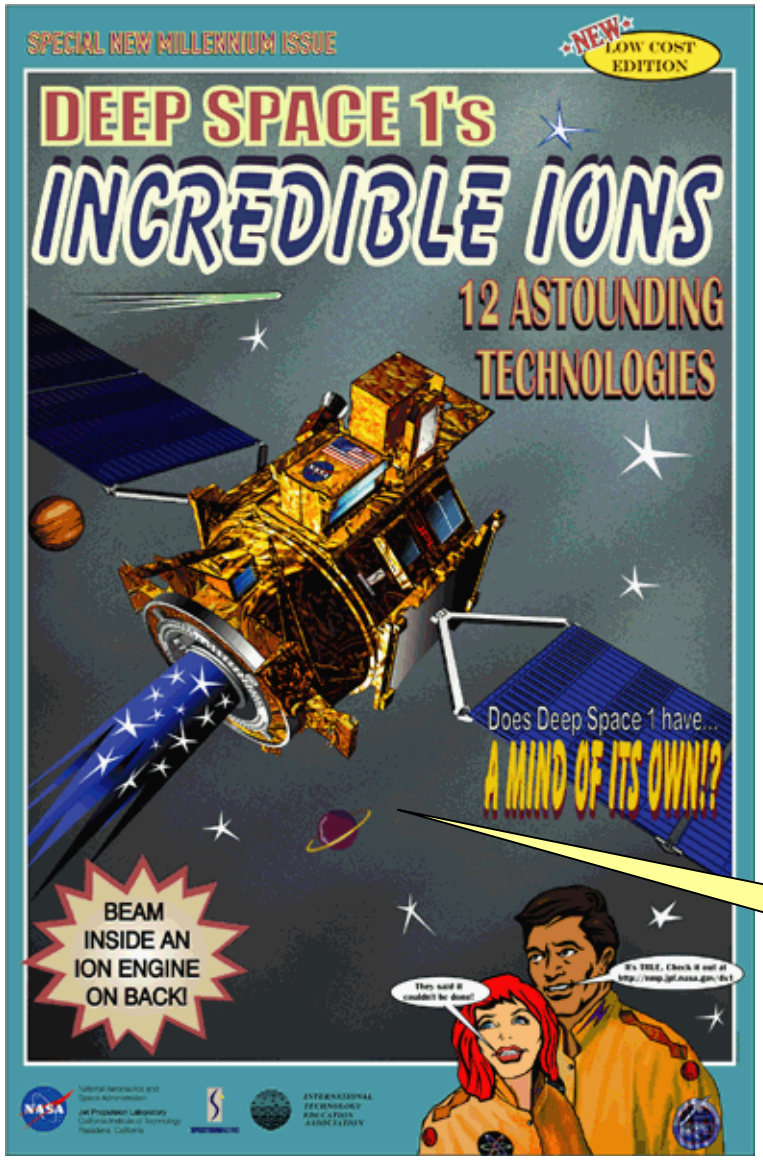
```
(eventOccursAt increase117086 cylinder117045)
+ (isa increase117086 IncreaseEvent)
+ (objectActedOn increase117086 pressure116998)
+ 19 other assertions
```

Example: Link analysis

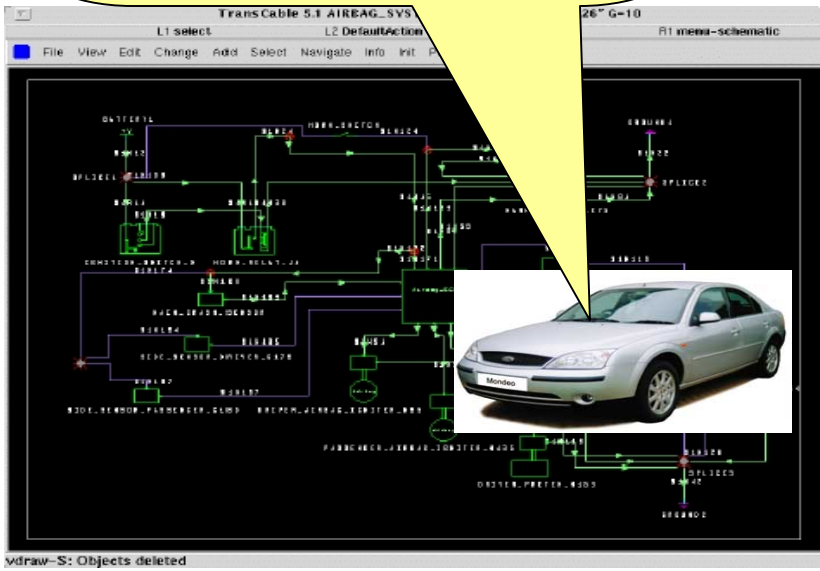


- Source: i2's Analyst's Notebook

Example: Model-based engineering

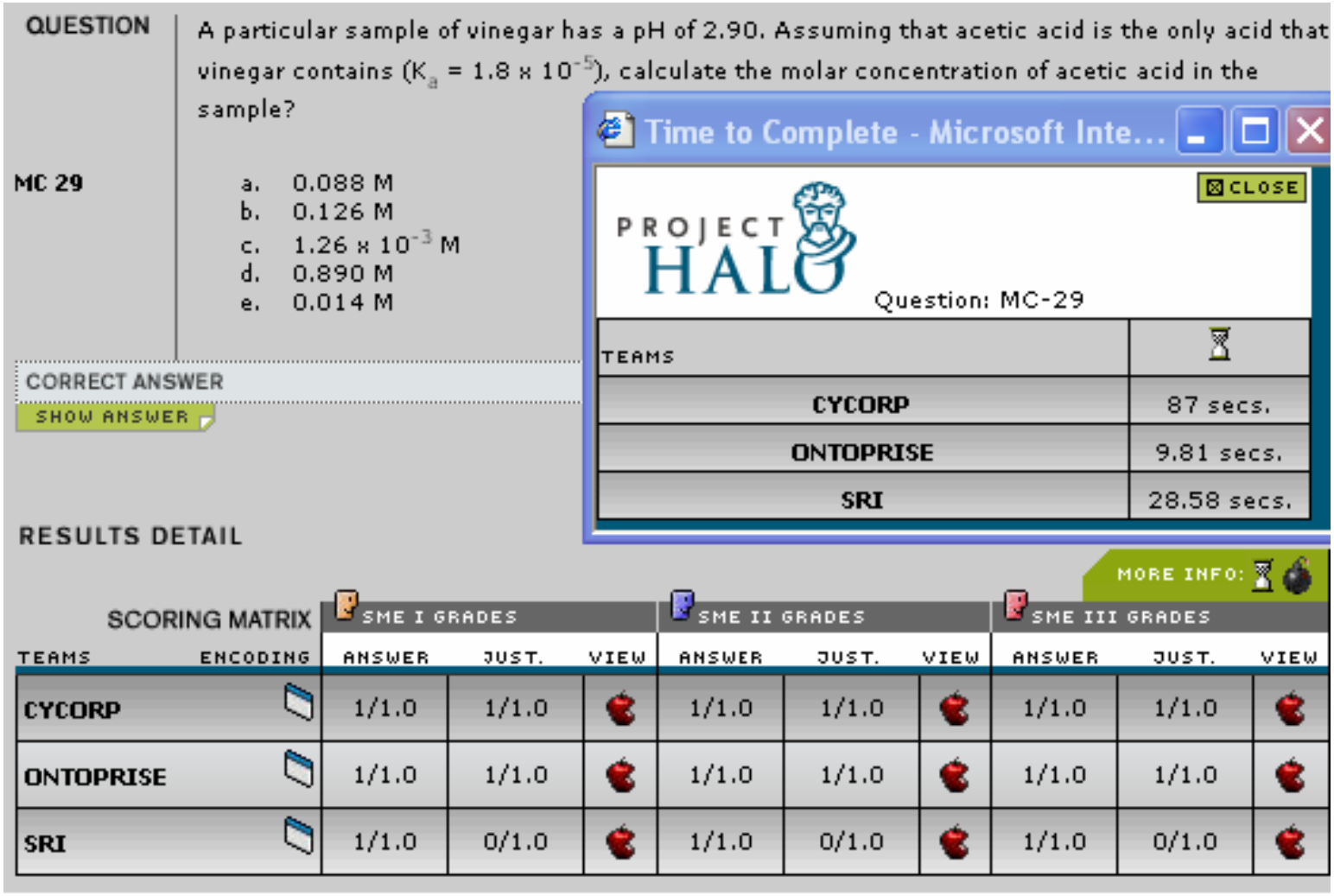


High-end cars have > 100 CPU's, diagnosis moving on-board



DS1 used model of its structure and function

Example: Scientific reasoning systems



QUESTION

A particular sample of vinegar has a pH of 2.90. Assuming that acetic acid is the only acid that vinegar contains ($K_a = 1.8 \times 10^{-5}$), calculate the molar concentration of acetic acid in the sample?

MC 29

a. 0.088 M
b. 0.126 M
c. 1.26×10^{-3} M
d. 0.890 M
e. 0.014 M

CORRECT ANSWER

SHOW ANSWER

Time to Complete - Microsoft Inte...

CLOSE

PROJECT HALO

Question: MC-29

TEAMS	
CYCORP	87 secs.
ONTOPRISE	9.81 secs.
SRI	28.58 secs.

RESULTS DETAIL

MORE INFO:

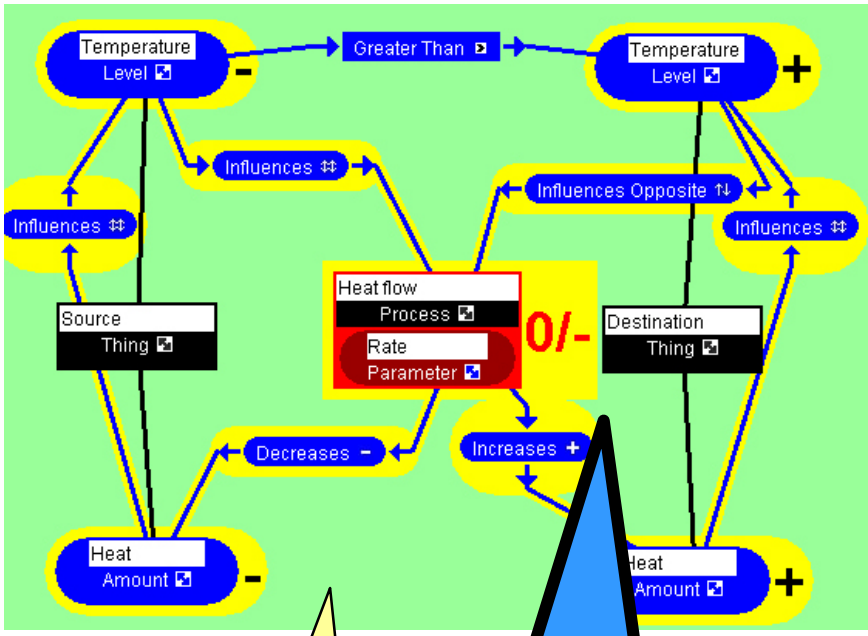
SCORING MATRIX		SME I GRADES			SME II GRADES			SME III GRADES		
TEAMS	ENCODING	ANSWER	JUST.	VIEW	ANSWER	JUST.	VIEW	ANSWER	JUST.	VIEW
CYCORP		1/1.0	1/1.0		1/1.0	1/1.0		1/1.0	1/1.0	
ONTOPRISE		1/1.0	1/1.0		1/1.0	1/1.0		1/1.0	1/1.0	
SRI		1/1.0	0/1.0		1/1.0	0/1.0		1/1.0	0/1.0	

- Source: www.projecthalo.com

Example: Intelligent tutoring systems

1. **Student:** angles are equal
Classified as: CONGRUENT-ANGLES
2. **Tutor:** Are any two angles congruent? What tells you that
there are two congruent angles here?
3. **Student:** angles in a triangle are equal
Classified as: CONGRUENT-ANGLES-IN-TRIANGLE
4. **Tutor:** That is a bit imprecise. Are ALL angles in ALL
triangles congruent?
5. **Student:** angles in an isosceles triangle are equal
Classified as: CONGRUENT-ANGLES-IN-ISOSCELESTRIANGLE
6. **Tutor:** Correct, except for one thing: are all three angles in
an isosceles triangle congruent?
7. **Student:** base angles in an isosceles triangle are equal
Classified as: ISOSCELES-TRIANGLE-BASE-ANGLES
8. **Tutor:** CORRECT.

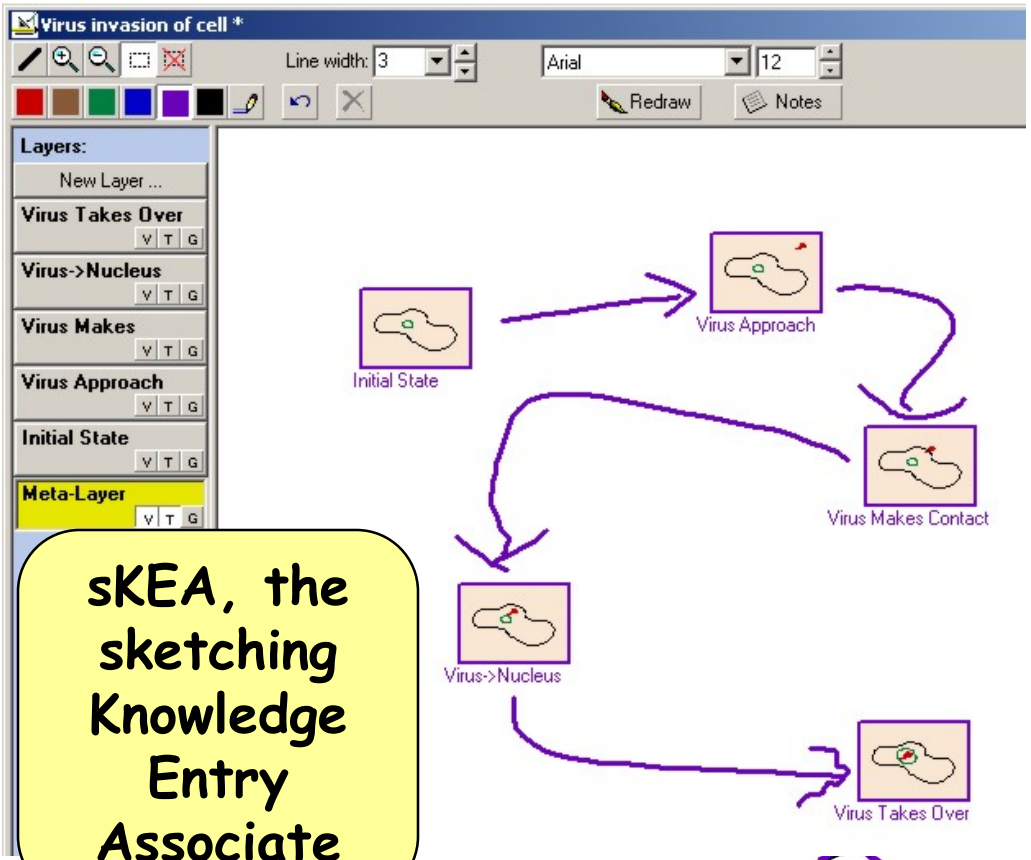
From Aleven, Popescu, and Koedinger 2001



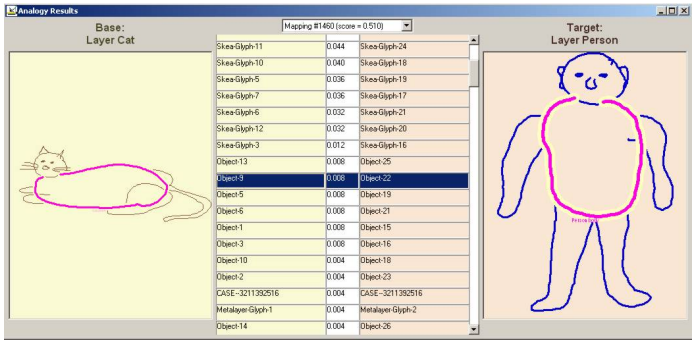
From NWU's VModel system for middle-school students

You predicted that Heat flow would be CONSTANT but instead it is DECREASING

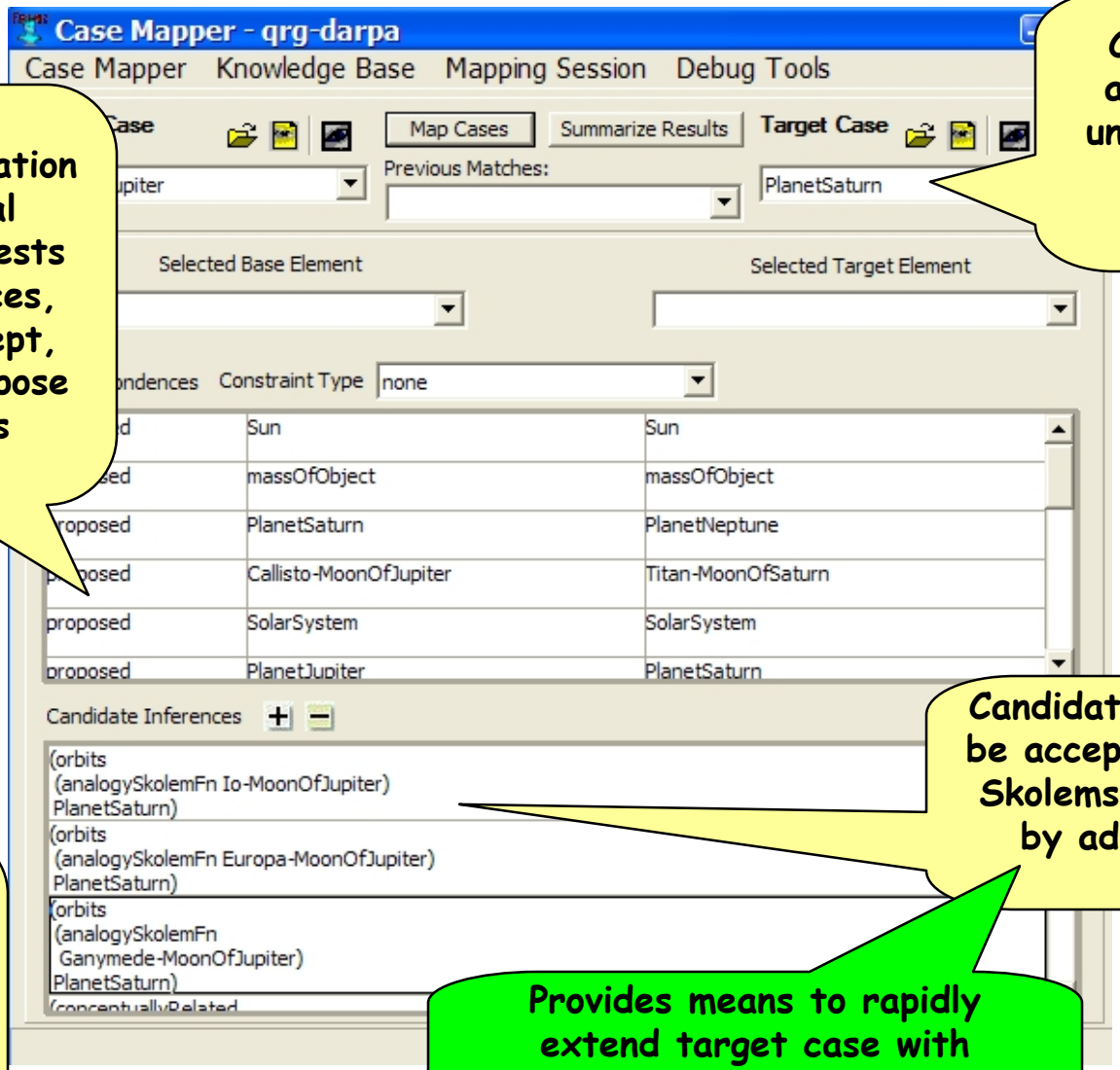
Example: Computers you can sketch with



sKEA, the sketching Knowledge Entry Associate



Example: Working through analogies



Cognitive simulation of analogical matching suggests correspondences, User can accept, reject, or propose alternatives

Cases constructed automatically from underlying knowledge base, using task constraints

Candidate inferences can be accepted or rejected. Skolems can be filled in by adding/selecting entities

Provides means to rapidly extend target case with knowledge imported from the base

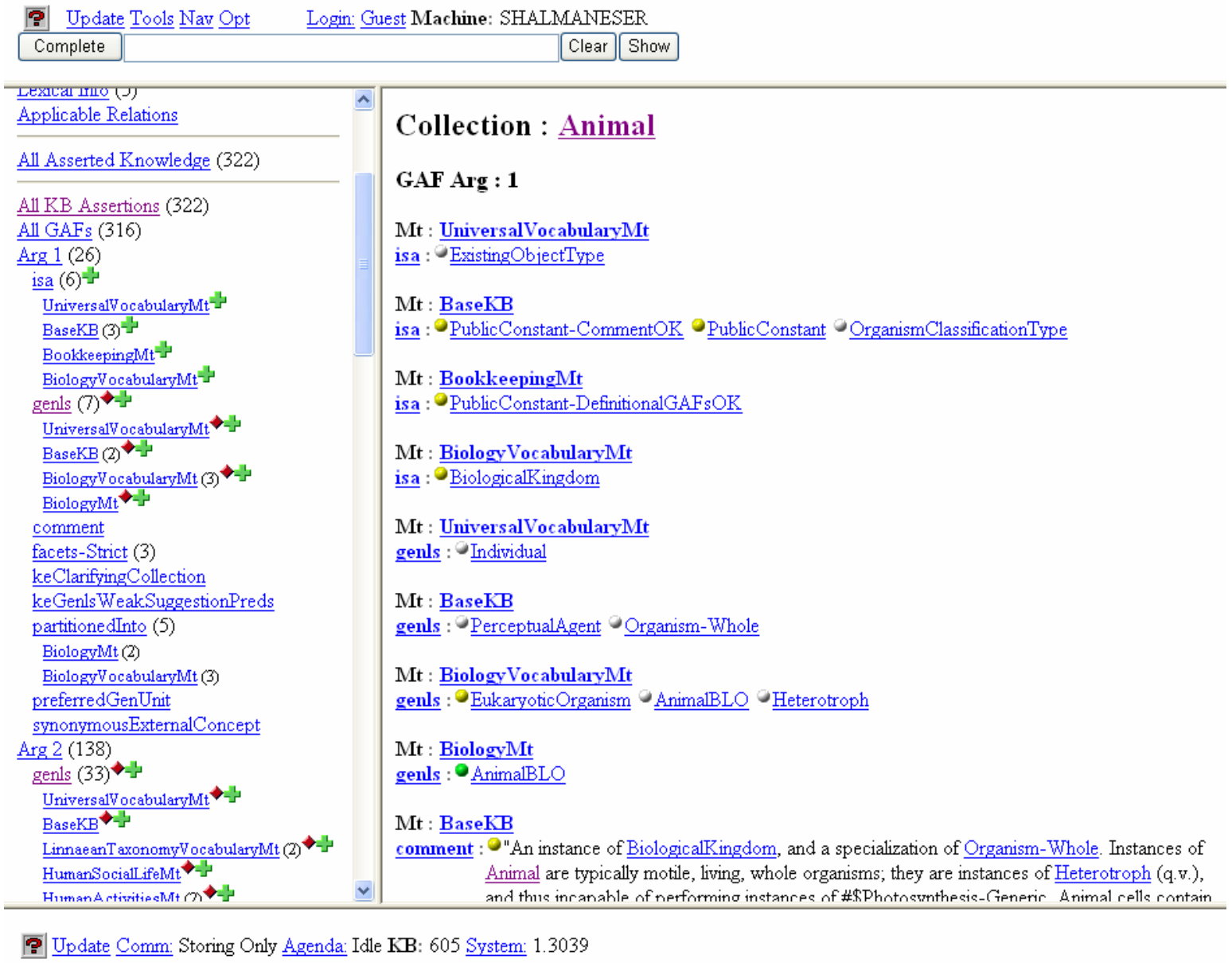
Case Mapper interface uses SME and MAC/FAC

Question: Why are you taking this class?

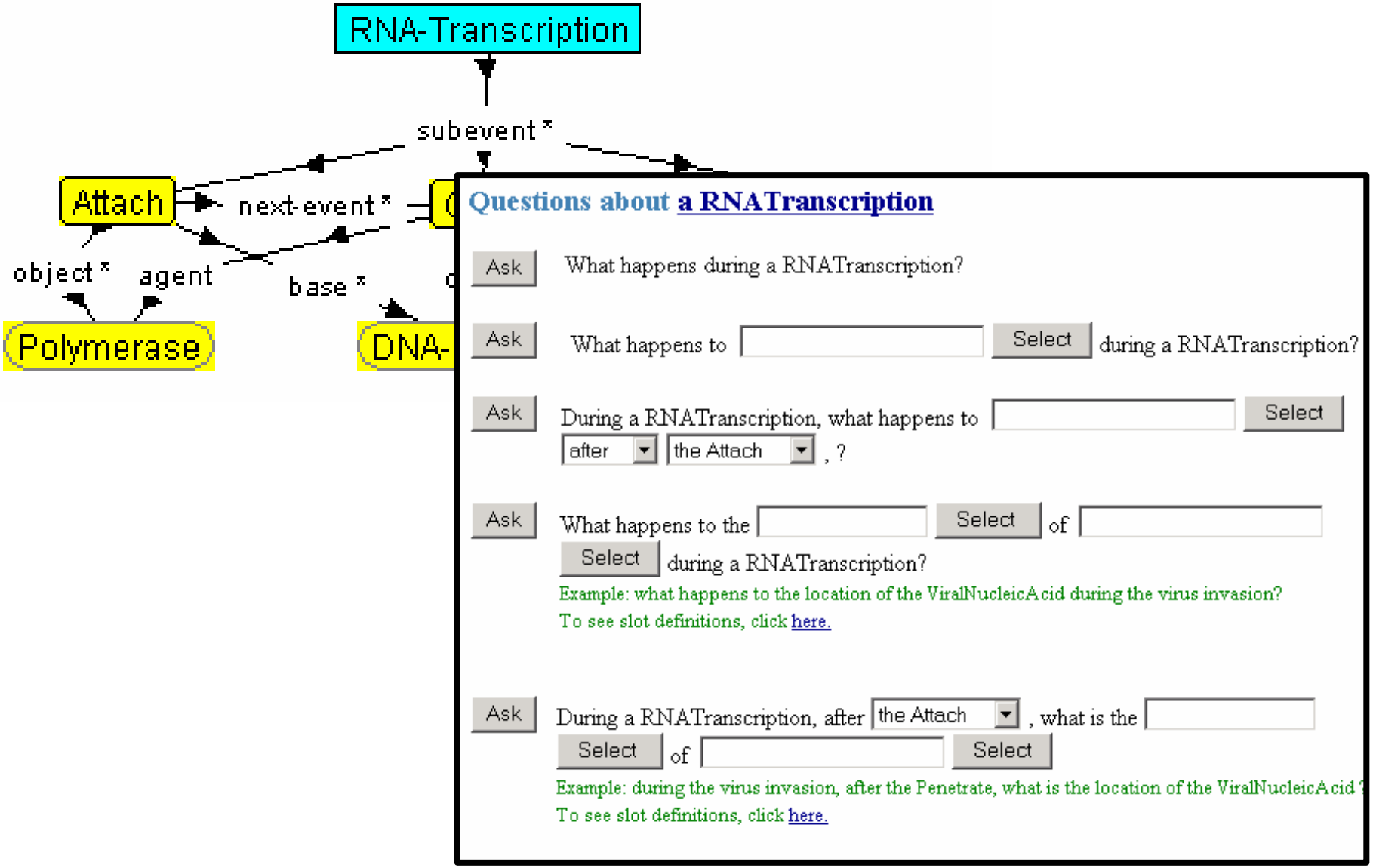
A hands-on approach

- We will be using pre-existing knowledge systems as tools
  - OpenCyc, from Cycorp
  - Shaken, from the SRI RKF team
- You will not be programming as part of this class
  - This class is not concerned with building/improving reasoning engines (see CS 344 for that!)
- You will be building representations using off-the-shelf tools and knowledge bases
  - They provide crucial support for dealing with large knowledge bases
  - They provide some built-in reasoning for testing your work

OpenCyc from Cycorp



SHAKEN from SRI RKF Team



Why two tools?

- Each has their strengths and weaknesses
  - As software
  - In their representation choices
- Comparing and contrasting them will provide more insight than looking at either alone

Experiment: Studio instruction

- How many laptops available?
- If not enough, we'll rely more on treeware

Textbook and other materials

- Brachman, R. and Levesque, H. (forthcoming) *Knowledge Representation and Reasoning*
  - We're using a "pre-release" version
    - Quartet Copies, 818 Clark Street, 847-328-0720
  - Provides excellent coverage of general principles, many interesting examples
- OpenCyc documentation
  - Very detailed and informative
  - http://www.opencyc.org/doc
- Shaken documentation and papers
  - http://www.ai.sri.com/project/SHAKEN

Syllabus

- Week 1: Introduction
- Week 2: Collections and KB organization
- Week 3: Events and their structure
- Week 4: Time and causality
- Week 5: Plans and goals
- Week 6: Quantity, values, behaviors
- Week 7: Qualitative mathematics and mechanisms
- Week 8: Knowledge and belief
- Week 9: TBD
- Rest of the time: In-class representation exercises and brainstorming, plus project work

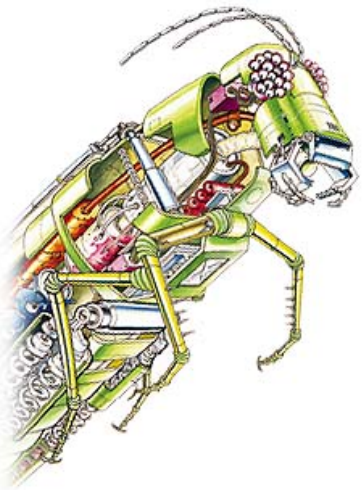
This will probably evolve over the quarter

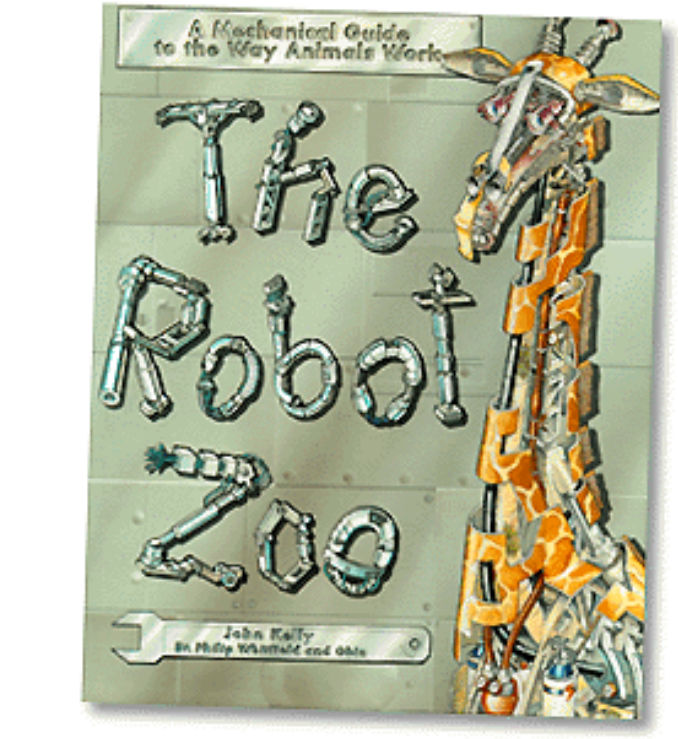
Grading

- Homework assignments
  - Most assignments will use one of these tools
  - Once project work starts in earnest, weekly progress reports
  - All work to be turned in via email to staff alias
- Class participation
  - Both in class and in on-line forums
  - Kicking around ideas together is key to learning
- Projects
  - Not programming projects or term papers
  - Domain theories and knowledge bases, plus a writeup of your design decisions and experiences
  - You'll use one of these tools in developing and testing your work

Projects

- Ideally, related to something you're interested in
  - e.g., thesis work, if you're a graduate student
  - e.g., Some hobby or passion, if undergraduate
- If you can't think of something, here are some ideas to get you started
  - We have extra resources available for projects in these areas



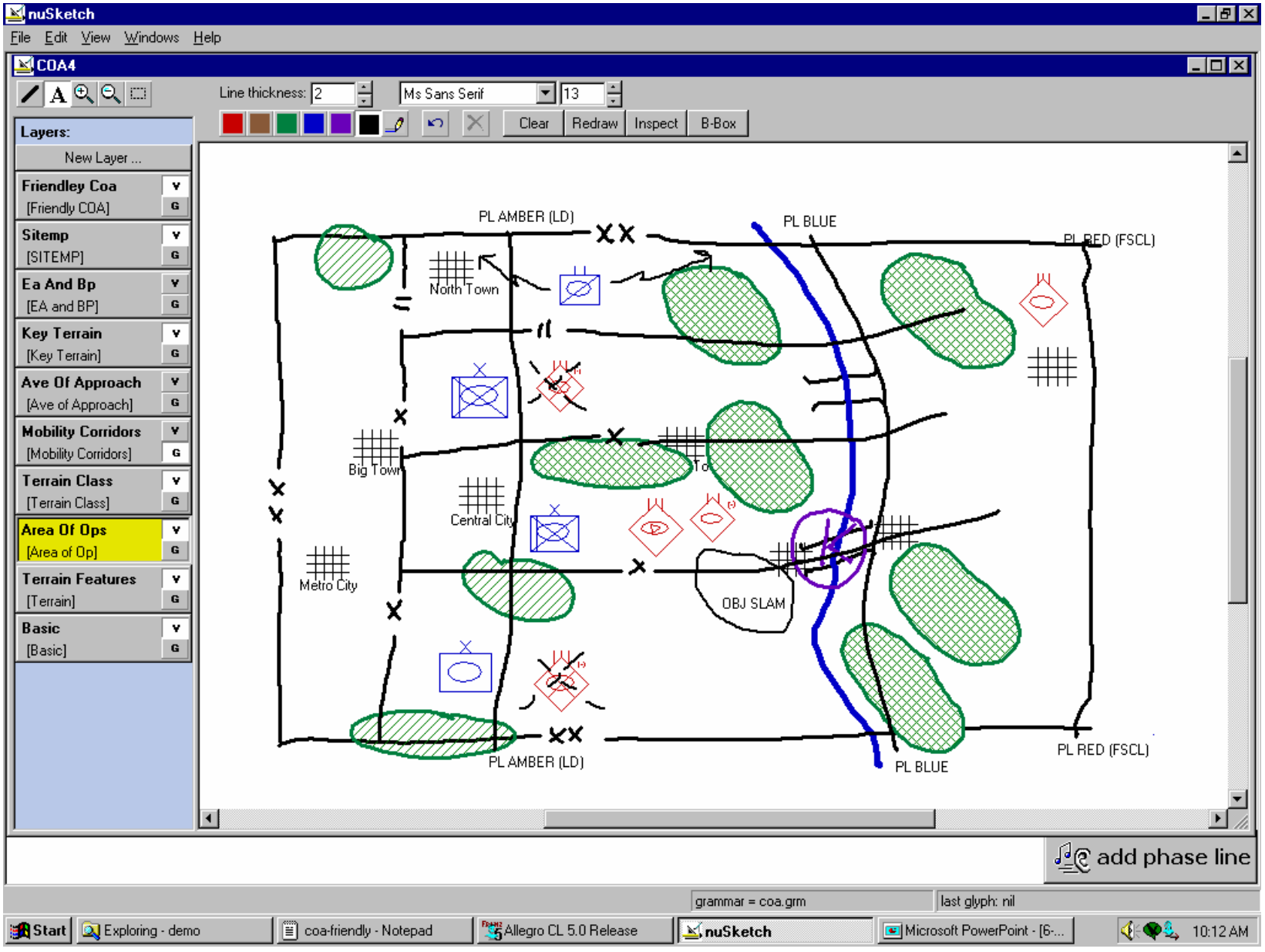


- Naïve biophysics
  - Structure and function of the parts of animals
  - Technological/biological analogies

Solar Energy

- How houses work
- What the sun does
- Different kinds of solar heating

Reasoning about Tactics



Qualitative economics and international relations

- ***Assuming constant production by Iran and others, would a 5% increase in production by Saudi Arabia have a positive or negative effect on the economy of Iran? Would a 5% increase by Kuwait have as large an effect?***
  - **Answer:** Negative. The Kuwaiti increase would not have as large an effect because Kuwait does not have as large a current production.

- ***What might the likely effects be on the current price of oil on the international market as a result of the terrorist attacks on Saudi oil facilities?***
  - **Answer:** Prices will increase dramatically in the short term due to supply shock and instability. Long term prices will normalize depending upon the severity of the damage to Saudi production, the ability of other suppliers to make up the supply shortage, and the termination of instability in the Persian Gulf.

Natural Language Semantics

- Connecting qualitative physics to verb semantics
- Connecting spatial/diagrammatic representations to spatial language
- Representing dialogue acts

Plausible reasoning about weather

- WHY project: Effort by Collins & Stevens in 1970s to build intelligent tutor for weather and climate
  - Weren't able to build a system for a variety of reasons, including
    - state of KR, NLP, QR, and analogical reasoning
    - Feeble computational resources available then
  - We could do much better today.

Your first assignment

- Install OpenCyc
  - Make sure you can
    - Browse the knowledge base
    - Make queries
    - Add new knowledge
    - Save what you did, and reload it later
- Reading:
  - Cyc 101 Tutorial, through “Errors in Representing Knowledge”
  - Chapters 1-3 in Brachman and Levesque