



# Overview of Cyc

Tom O'Hara  
NMSU AI Seminar  
18 February 2002

# History

- ⇒ AM and Eurisko in early 1970's
- ⇒ Japanese 5th Generation Project in early 1980's
  - expert systems & neural networks
- ⇒ Microelectronics and Computer Technology Company (MCC) founded in 1984
  - Cyc project & nine others
- ⇒ Cycorp founded in 1995
- ⇒ OpenCyc announced 2001

# Cyc Knowledge Base Overview

- ⇒ Very large knowledge base
  - 100,000+ terms
  - 1,000,000+ assertions
- ⇒ CycL representation language
- ⇒ Microtheories for structuring the KB
- ⇒ Lexicon
  - mapping from English to CycL
  - not covered in talk

# Cycl representation language

- ⇒ Problems with frame-based representation
  - difficulty representing assertions with arity higher than 2
  - quantification not directly expressible
  - difficulty in representing meta-assertions
- ⇒ Cycl based on First Order Predicate Logic (FOPL)
  - extension to KIF (Knowledge Interchange Format)

# Details of Cycl

## ⇒ Constants

- Case-sensitive concept identifier
- examples: Cyc, DougLenat, BaseKB, EnglishWord

## ⇒ Variables

- Case-insensitive identifier starting with ? symbol
- examples: ?TYPE

## ⇒ Predicates

- Terms that represent relation types defined in the KB
- examples: isa, genls, comment

# More Details of Cycl

## ⇒ Formulas

- An expression of the form (*predicate arg1 arg2 ...*)
- Truth value: {true, default true, false, default false, unknown}
- Examples:
  - (isa Dog BiologicalSpecies)
  - (genls Dog Carnivore)
  - (skillCapableOf LinusVanPelt PlayingAMusicalInstrument performedBy)

## ⇒ Logical connectors

- examples: not, and, or, implies

## ⇒ Quantifiers

- examples: forAll, thereExists

# Final Details of Cycl

⇒ Rule macro predicates

```
(forall ?A
  (implies
    (isa ?A Animal)
    (thereExists ?M
      (mother ?A ?M))))
```

replaced with

```
(relationAllExists mother Animal Animal)
```

⇒ Non-atomic terms

- Functional terms: (FruitFn AppleTree)
- Reifiable versus non-reifiable functions

# Important features of the KB

⇒ Hierarchy with two dominance relations

- isa for instance type specification

(isa DougLenat HumanCyclist)

- genls for type generalization

(genls HumanCyclist Human)

⇒ Individuals versus Collections

- Allows fine distinctions in assertions

- But complicates knowledge engineering

# Cyc Inference Engine

- ⇒ Proprietary algorithm
- ⇒ Epistemological Level (EL) vs. Heuristic Level (HL)
- ⇒ HL Modules
  - special purpose inferencing
  - interface for defining new modules
- ⇒ Inferencing no longer complete

# Applications of Cyc

- ⇒ Database Integration
- ⇒ HPKB: High-Performance Knowledge Bases
- ⇒ e-Cyc: Web searching
- ⇒ RKF: Rapid Knowledge Formation
- ⇒ AQUAINT: Question Answering

# Upper Cye Ontology

- ⇒ subset of KB available for downloading
- ⇒ approximately 3,000 terms & 13,000 assertions
- ⇒ "general concepts of human consensus reality"

# Predicate usage in Upper Cyc

| <i>Freq.</i> | <i>Predicate</i> | <i>Description</i>                          |
|--------------|------------------|---|
| 4503         | isa              | instance of type                            |
| 2695         | comment          | comment describing term usage               |
| 2565         | genls            | type generalization                         |
| 920          | arg1Isa          | argument 1 constraint                       |
| 836          | arg2Isa          | argument 2 constraint                       |
| 525          | genlPreds        | predicate generalization                    |
| 301          | not              | logical not connective                      |
| 243          | resultIsa        | function result type                        |
| 120          | arg3Isa          | argument 3 constraint                       |
| 107          | implies          | logical implication (i.e., rule definition) |

# Pro's and Con's of Cyc

## ⇒ Kudos

- Chosen as standard for HPKB follow-up work
- Fairing well in current RKF project (IET 2001)
- Cyc project still active after nearly two decades of work

## ⇒ Criticisms

- Common Knowledge or Superior Ignorance? (Locke 1990)
- Promising but not yet suitable for NLP (Mahesh et al. 1996)
- Promising but not readily usable at IRS (Sanguino 2001)

# Bibliography

IET (2001), "RKF Y1 Evaluation Report", October 2001,  
<http://www.iet.com/Projects/RKF/IET-RKF-Y1-Evaluation.ppt>.

Lenat, D. B. and R. V. Guha (1990), *Building Large Knowledge Based Systems*. Reading, Massachusetts: Addison Wesley.

Lenat, D. B. (1995), "Cyc: A Large-Scale Investment in Knowledge Infrastructure." *Communications of the ACM* 38, no. 11.

Mahesh, K., S. Nirenburg and S. Beale (1996), "KR Requirements for Natural Language Semantics: A Critical Evaluation of Cyc". *Proceedings of KR-96*.

Sanguino, Roland (2001), "Evaluation of Cyc ", LEF grant report, CSC, Miami, FL, March 2001, [http://www2.csc.com/lef/programs/grants/finalpapers/sanguino\\_eval\\_cyc.pdf](http://www2.csc.com/lef/programs/grants/finalpapers/sanguino_eval_cyc.pdf).

Russell , Stuart and Peter Norvig (1995), *Artificial Intelligence: A Modern Approach*, Upper Saddle River, NJ: Prentice-Hall.

Whitten, David (1997), The Unofficial, Unauthorized Cyc Frequently Asked Questions Information Sheet, <http://www.robotwisdom.com/ai/cycfaq.html>.