

# OpenCyc Commonsense AI Tutorial

1

**STEPHEN L. REED**  
**PRINCIPAL DEVELOPER**

**TEXAI.ORG**

**Austin, Texas**

# History of the Cyc Project

2

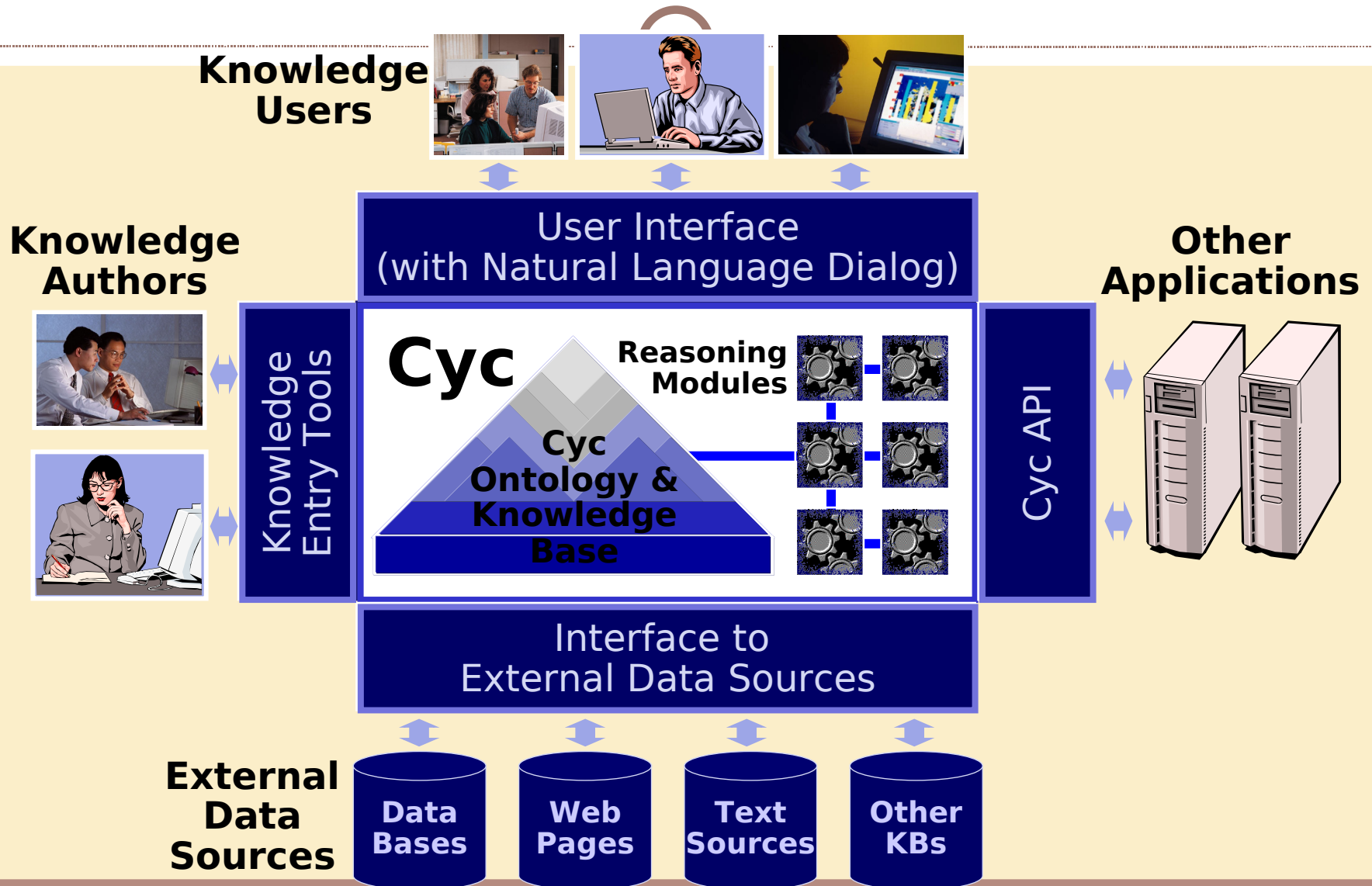
- 1982, Japan begins the Fifth Generation Computer project
- 1982, the Microelectronics and Computer Technology Corp (MCC) formed in response
- 1984 MCC recruits Doug Lenat from Stanford to create a commonsense knowledge base overcoming the brittleness of then current expert systems
- 1995 - As corporate sponsorships diminished, the Cyc project was spun-off into Cycorp, the company. The name is a play on an entity in the Babylon 5 TV series
- Late 1990's – Cycorp completes the tenth rewrite of its inference engine and object store and begins its migration from Symbolics Lisp Machines to Linux PCs
- 1999 Cycorp strengthens its relationships with US military and intelligence community sponsors, which to this day provide the majority of its funding – believing that a commonsense ontology is a hub for integrating disparate military and intelligence systems
- 2001 – OpenCyc released
- 2006 – Cycorp implements a Java runtime for its Lisp source code

# What is OpenCyc?

3

- Developed by Cycorp, a government sponsored research company in Austin, Texas
- A free, comprehensive ontology
- Hundreds of thousands of terms, mostly classes of things
- Over a million logical statements defining those terms
- Manually created by a team of philosophers over a 20 year duration
- A large portion is compatible with Resource Description Framework (RDF) the logical language of the Semantic Web
- Hosted at SourceForge
- Temporarily offline while the proprietary object store and inference engine is converted to Java
- Cycorp published OpenCyc in order to promote its ontology as a standard for the semantic web

# Cyc Reasoning System



# What Gap in AGI Does OpenCyc Fill?

5

- A world model is required by any general problem-solving AI
- Concepts in the world model range in level of abstraction from:
  - the sub-symbolic, e.g. perceived sound wave
  - Symbolic, e.g. name and address
- OpenCyc provides a candidate schema for a comprehensive symbolic world model
- An essential aspect of general problem-solving is the use of inference – acting upon conclusions derived from observations, experience or premises
- OpenCyc's knowledge representation format is symbolic, and is designed for deductive inference, and has been demonstrated with planning, induction, and abduction as well
- A recursively self-improving AGI should take advantage of existing structured knowledge sources
- OpenCyc technology has been demonstrated to integrate structured knowledge sources



# OpenCyc Fundamentals

7

- Symbolic concepts are represented as atomic terms, e.g. `TransportationDevice`, or a composed terms, e.g. `(FruitFn AppleTree)`
- Besides concepts, terms can also be literals, e.g. `true`, `false`, `"abcdef"`, `1.0`.
- Relations about terms are represented by assertions, having a named predicate and from one to six argument positions filled by a term or another assertion
- Each assertion is placed in a context, called a microtheory. These contexts are also terms and are arranged by OpenCyc into a generalization hierarchy for inference
- Categories of OpenCyc concepts
  - Collection (RDF Class)
  - Individual (RDF Individual)
  - Predicate (RDF Property)
  - Microtheory (RDF named graph)
- Single rooted concept hierarchy with *Thing* at the top
- Multiple inheritance
- Class cross-cutting aspects
  - Temporal vs non-temporal
  - Object-like vs Stuff-like
  - Partially-tangible vs intangible
  - Individual vs SetOrCollection
  - Whether something is a situation or not

# A Tour of OpenCyc

8

- OpenCyc.org [here](#)
  - A lot of documentation and tutorials
  - Cyc 101 – periodic Cyc classes in Austin
- Hosted at SourceForge
- Doug Lenat's presentation at Google [here](#)
- Knowledge Base browser [here](#)
  - Definitional assertions
  - Hierarchy browser
- Cyc Foundation OpenCyc endpoint [here](#)
- Cyc vocabulary introduction (e.g. event actors) [here](#)



# Linking Open Data

9

- Databases and other structured knowledge sources expose their contents to the World Wide Web
- Each exposed concept is identified with a URI, i.e. web address
- The address returns something descriptive to humans when viewed with a browser
- Relationships between concepts are expressed as RDF statements: subject, predicate (i.e. property), object, and optional context
- OWL and RDF provide standard schema defining properties: `rdf:type`, `rdfs:subClassOf`, `owl:sameAs`, etc.
- Linked Open Data forms part of the Semantic Web infrastructure. Now need to construct the intelligent agents that use the linked data.
- OpenCyc is a contender for the standard ontology for linking open data
- The UMBEL topic ontology was derived from OpenCyc and has 20000 class terms
- Notable LOD datasets:
  - DBpedia – derived from Wikipedia
  - YAGO – 20 million facts about 2 million named entities
  - GeoNames – 6.5 million facts about geographical locations



# “Learn proper nouns” → logic



Assignment-Obligation2 **rdf:type** cyc:Assignment-Obligation .

Assignment-Obligation2 **cyc:allottedAgents** Addressee .

Assignment-Obligation2 **cyc:assigner** Speaker .

Assignment-Obligation2 assignmentPostCondition Learning3 .

( you )

Learning3 typeOrSubClassOf **cyc:Learning** .

Learning3 **cyc:actionFulfillsAssignment** Assignment-Obligation2 .

Learning3 **cyc:situationConstituents** Addressee .

Learning3 **cyc:performedBy** Addressee .

Learning3 **cyc:thingComprehended** ProperCountNoun1 .

Learning3 fcgDiscourseRole Addressee .

Learning3 fcgStatus SingleObject .

Learning3 situationHappeningOnDate **cyc:Now** .

learn

ProperCountNoun1 typeOrSubClassOf **cyc:ProperCountNoun** .

ProperCountNoun1 fcgDiscourseRole External .

ProperCountNoun1 fcgStatus MultipleObjects .

proper nouns

# 'BethLynn Maxwell is a proper noun → logic



LexicalWord1 **rdf:type** FCGClauseSubject .

LexicalWord1 typeOrSubClassOf **cyc:LexicalWord** .

LexicalWord1 **cyc:wordStrings** "BethLynn Maxwell" .

LexicalWord1 fcgDiscourseRole External .

LexicalWord1 fcgStatus SingleObject .

'Bethlynn'

ImplicationSituation3 typeOrSubClassOf ImplicationSituation .

ImplicationSituation3 **cyc:situationConstituents** LexicalWord1 .

ImplicationSituation3 implicationAntecedant LexicalWord1 .

ImplicationSituation3 implicationConsequent ProperCountNoun2 .

ImplicationSituation3 fcgDiscourseRole External .

ImplicationSituation3 fcgStatus SingleObject .

ImplicationSituation3 situationHappeningOnDate **cyc:Now** .

is

ProperCountNoun2 typeOrSubClassOf IndefiniteThingInThisDiscourse .

ProperCountNoun2 typeOrSubClassOf **cyc:ProperCountNoun** .

ProperCountNoun2 fcgDiscourseRole External .

ProperCountNoun2 fcgStatus SingleObject .

a proper  
noun

# SPARQL that matches “X is a proper noun”



```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX cyc: <http://sw.cyc.com/2006/07/27/cyc/>
PREFIX texai: <http://texai.org/texai/>
```

```
SELECT ?LexicalWord1 ?CharacterString
WHERE {
  ?LexicalWord1 cyc:wordStrings ?CharacterString .
  ?LexicalWord1 rdf:type cyc:LexicalWord .
  ?ProperCountNoun2 rdf:type cyc:ProperCountNoun .
  _:ImplicationSituation3 texai:implicationAntecedant ?LexicalWord1 .
  _:ImplicationSituation3 texai:implicationConsequent ?ProperCountNoun2 .
}
```

- Texai uses this query to perceive that the character string ‘BethLynn’ is to be used when creating the morphological rule for the corresponding proper noun

# OpenCyc Annoyances

14

- Does not include the several million commonsense assertions and rules from full proprietary Cyc
  - ResearchCyc does that
- OpenCyc's inference engine and object store are not free software
- No current provision for vetting and incorporating volunteer ontology input
- Continues as an incompatible superset of the Semantic Web (W3C) RDF/OWL standard
- Authored by philosophers and mathematicians, thus an impedance mismatch with the needs of computational linguists
- Fine cross-cutting distinctions in the upper ontology make integration of lower level concepts more difficult – i.e. disjointness is rampant
- Due to staff turnover, the passage of time, initially poor authoring guidelines, and until a few years ago – lack of unit tests, quality (e.g. well-formedness) and coverage is inconsistent
  - Cyc authored what its sponsors funded
- Often there are multiple approaches to encoding the same knowledge
  - Davidsonian events vs direct assertions between role players in an event

# Summary & Questions



- At some point in AGI development, it will be useful to incorporate OpenCyc content
- Likewise, Linked Open Data, mapped with a shared ontology, is a useful knowledge source input to an AGI, or conversely a means by which an AGI can disseminate its own knowledge.

Questions???

And enjoy the rest of AGI-09!