Ontology Archeology: Mining a Decade of Effort on the Suggested Upper Ontology

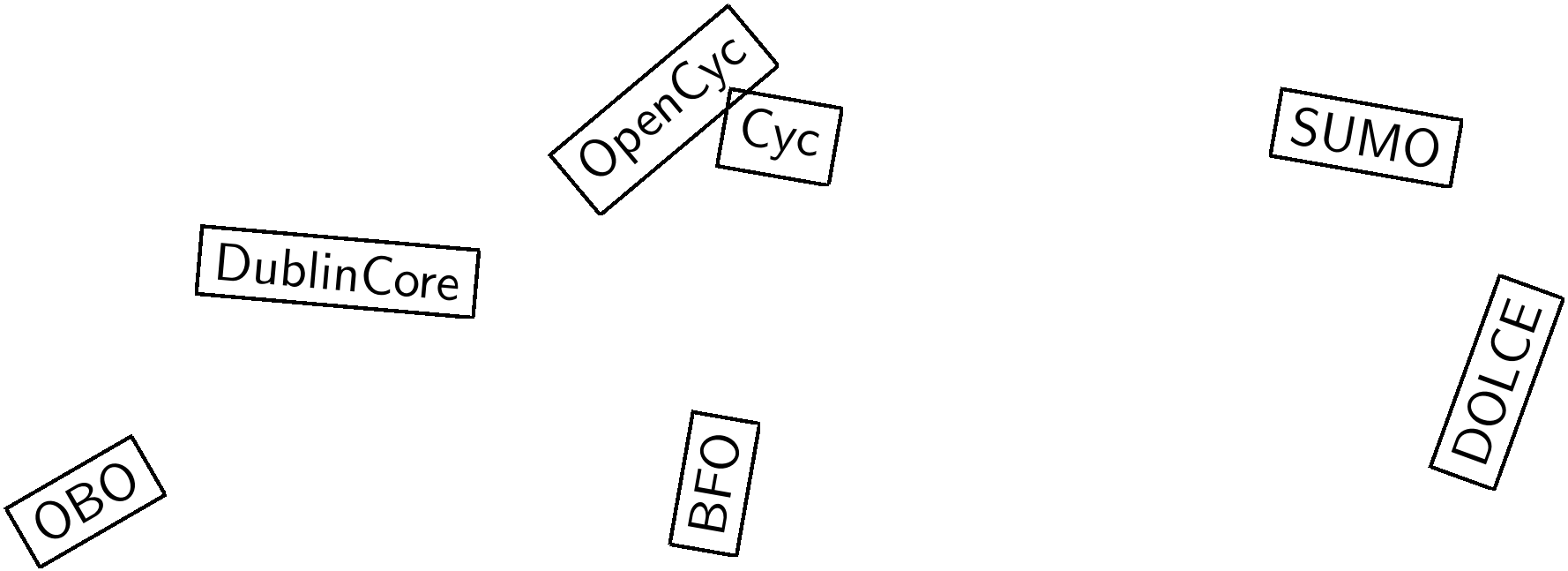
Adam Pease and Christoph Benzmüller*

Articulate Software, Angwin, CA, USA

ARCOE-10, August 16-17, 2010 Lisbon, Portugal

*supported by DFG grant BE 2501/6-1

Ontology: An (Over-)Stretched Notion?



Ontology versus Taxonomy

Notion of Taxonomy

Characteristics

- simple hierarchical categorization and classification of entities in a domain (simple ontology in form of a hierarchy)
- classes, super-/subclasses, related classes, maybe some extras such as cardinality constraints, etc.
- computational properties $>$ expressivity

Analogy?

- taxonomy development $\approx$

GUI builder (restr. lang. features)

Notion of Taxonomy

Characteristics

- simple hierarchical categorization and classification of entities in a domain (simple ontology in form of a hierarchy)
- classes, super-/subclasses, related classes, maybe some extras such as cardinality constraints, etc.
- computational properties $>$ expressivity

Analogy?

- taxonomy development $\approx$

GUI builder (restr. lang. features)

Notion of Ontology

Characteristics

- rich semantic specification of a conceptualization; describes terms & entities and their relationships in a domain (often includes a taxonomy)
- expressive rules, e.g. first-order, sometimes even higher-order
- expressivity $>$ computational properties

Analogy?

- ontology development $\approx$ unrestricted, large-scale software development

Notion of Ontology

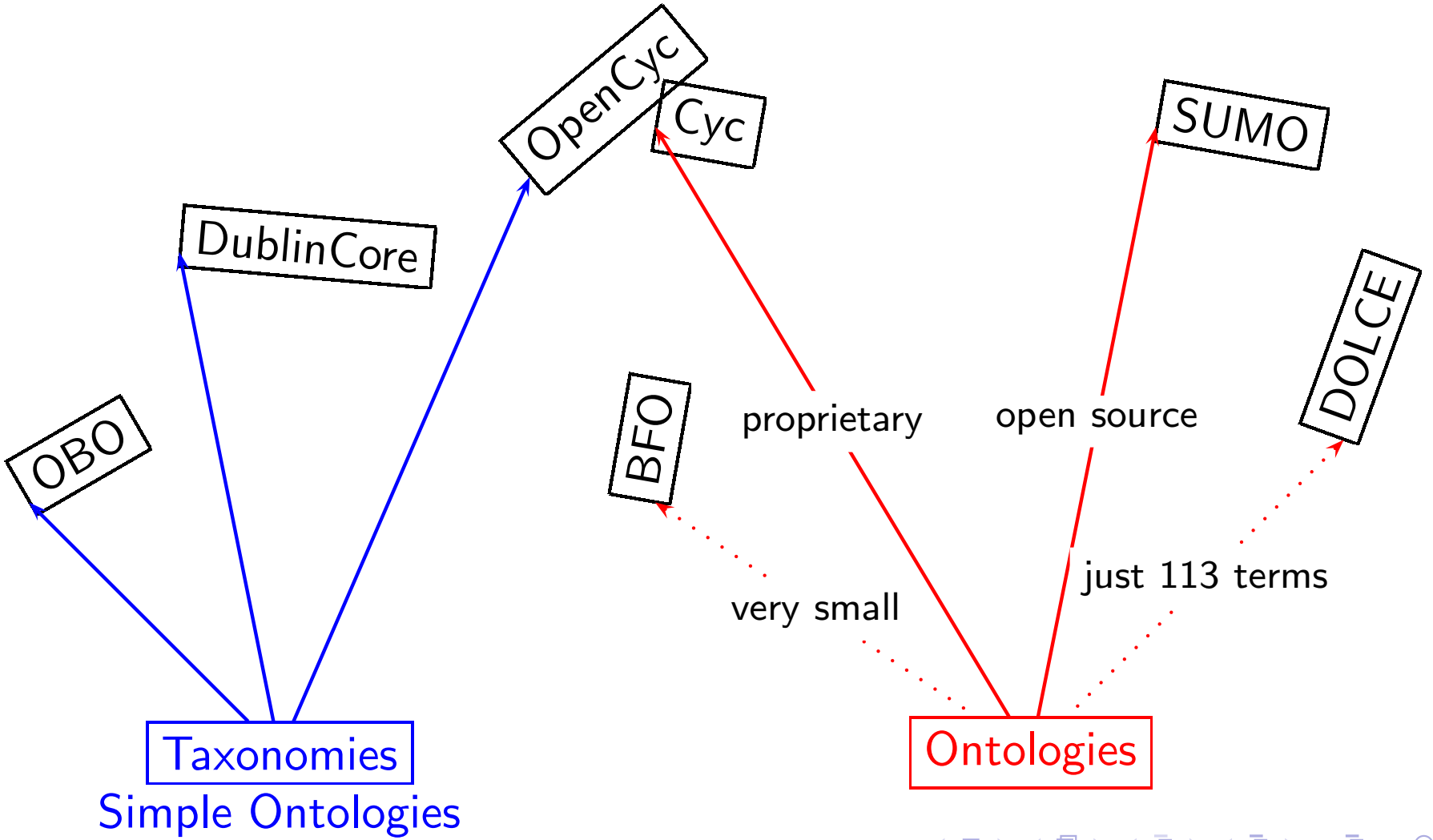
Characteristics

- rich semantic specification of a conceptualization; describes terms & entities and their relationships in a domain (often includes a taxonomy)
- expressive rules, e.g. first-order, sometimes even higher-order
- expressivity $>$ computational properties

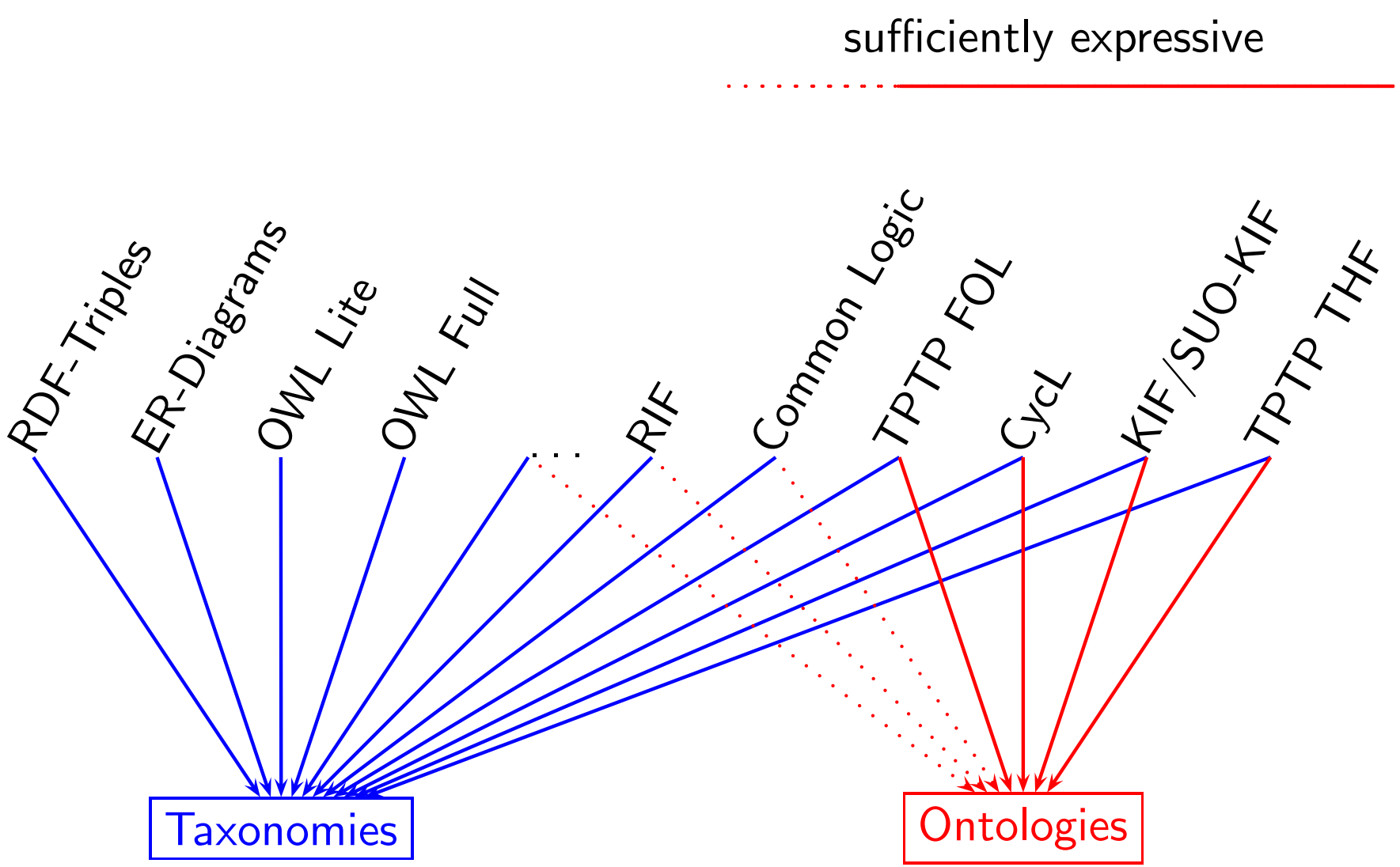
Analogy?

- ontology development $\approx$ unrestricted, large-scale software development

Ontology: An (Over-)Stretched Notion?



Ontology: An (Over-)Stretched Notion?



Taxonomy versus Ontology: Examples

Simple Taxonomy (excerpt from SUMO's base taxonomy)

```
(subclass Physical Entity)
(subclass Abstract Entity)
(partition Entity Physical Abstract)

(subclass Object Physical)
(subclass Process Physical)
(partition Physical Object Process)


...
```

Taxonomy versus Ontology: Examples

The SUMO Ontology is more than just a Taxonomy

```
(subclass Physical Entity)
(subclass Abstract Entity)
(partition Entity Physical Abstract)

(subclass Object Physical)
(subclass Process Physical)
(partition Physical Object Process)
```

```
(<=>
    (instance ?PHYS Physical)
    (exists (?LOC ?TIME)
        (and
            (located ?PHYS ?LOC)
            (time ?PHYS ?TIME))))
```

Taxonomy versus Ontology: Examples

The SUMO Ontology is more than just a Taxonomy

```
(subclass Physical Entity)
(subclass Abstract Entity)
(partition Entity Physical Abstract)

(subclass Object Physical)
(subclass Process Physical)
(partition Physical Object Process)
```

```
(<=>
    (instance ?PHYS Physical)
    (exists (?LOC ?TIME)
        (and
            (located ?PHYS ?LOC)
            (time ?PHYS ?TIME))))
```

```
(subclass ObjectAttitude IntentionalRelation)

(=>
    (and
        (instance ?REL ObjectAttitude)
        (?REL ?AGENT ?THING))
    (instance ?THING Physical))
```

More About SUMO

The SUMO Ontology: Some Facts

- developed since: $>$ 10 years
- original motivation: education support
  meta data for government training applications; did not really narrow the research focus though; observation in different specialist domains: need for common upper level ontology
- open source: `www.ontologyportal.org`
- SUMO versus SUMO:

  SUMO: Suggested Upper-level Ontology
  MILO: Mid-level Ontology
  Specific Domain-level Ontologies
  } SUMO

- representation language: SUO-KIF
  (adaptation of the Knowledge Interchange Format KIF)
- logic: mainly first-order; some higher-order extensions

The SUMO Ontology: Some Facts

- developed since: $>$ 10 years
- original motivation: education support
  meta data for government training applications; did not really narrow the research focus though; observation in different specialist domains: need for common upper level ontology
- open source: www.ontologyportal.org
- SUMO versus SUMO:

  SUMO: Suggested Upper-level Ontology
  MILO: Mid-level Ontology
  Specific Domain-level Ontologies
  } SUMO

- representation language: SUO-KIF
  (adaptation of the Knowledge Interchange Format KIF)
- logic: mainly first-order; some higher-order extensions

The SUMO Ontology: Some Facts

- developed since: $>$ 10 years
- original motivation: education support
  meta data for government training applications; did not really narrow the research focus though; observation in different specialist domains: need for common upper level ontology
- open source: `www.ontologyportal.org`
- SUMO versus SUMO:

  SUMO: Suggested Upper-level Ontology
  MILO: Mid-level Ontology
  Specific Domain-level Ontologies
  } SUMO

- representation language: SUO-KIF
  (adaptation of the Knowledge Interchange Format KIF)
- logic: mainly first-order; some higher-order extensions

The SUMO Ontology: Some Facts

- developed since: $>$ 10 years
- original motivation: education support
  meta data for government training applications; did not really narrow the research focus though; observation in different specialist domains: need for common upper level ontology
- open source: `www.ontologyportal.org`
- SUMO versus SUMO:

  SUMO: Suggested Upper-level Ontology
  MILO: Mid-level Ontology
  Specific Domain-level Ontologies
  } SUMO

- representation language: SUO-KIF
  (adaptation of the Knowledge Interchange Format KIF)
- logic: mainly first-order; some higher-order extensions

The SUMO Ontology: Some Facts

- developed since: $>$ 10 years
- original motivation: education support
  meta data for government training applications; did not really narrow the research focus though; observation in different specialist domains: need for common upper level ontology
- open source: `www.ontologyportal.org`
- SUMO versus SUMO:

  SUMO: Suggested Upper-level Ontology
  MILO: Mid-level Ontology
  Specific Domain-level Ontologies
  } SUMO

- representation language: SUO-KIF
  (adaptation of the Knowledge Interchange Format KIF)
- logic: mainly first-order; some higher-order extensions

The SUMO Ontology: More Facts

SUMO			
Total Terms	relations	Total Axioms	Rules
1173	353	4741	932
MILO			
Total Terms	relations	Total Axioms	Rules
1662	159	5116	1183
Domain ontologies			
Total Terms	relations	Total Axioms	Rules
17312	708	77974	2041
Total for all ontologies			
Total Terms	relations	Total Axioms	Rules
20147	1220	87831	4156

Table 1: SUMO term and axiom statistics

SUMO Example Entries

SUMO — 'part'

```
(domain part 1 Object)
(domain part 2 Object)
(instance part PartialOrderingRelation)
(instance part SpatialRelation)
...
```

MILO — 'typical-part'

```
(domainSubclass typicalPart 1 Object)
(domainSubclass typicalPart 2 Object)
(instance typicalPart BinaryRelation)
(instance typicalPart SpatialRelation)
(relatedInternalConcept typicalPart part)
...
```

Government Domain Ontology — 'capitalCity'

```
(domain capitalCity 1 City)
(domain capitalCity 2 GeopoliticalArea)
(instance capitalCity BinaryPredicate)
(subrelation capitalCity administrativeCenter)
...
```

How Developed?

SUMO

- initially by mining theories of common sense knowledge (e.g. James Allen's theory of temporal relations)
- by reflection and inspection of world; most of SUMO now original

MILO

- WordNet used to check coverage of SUMO/MILO
- criterion: every WordNet synset that occurred $\geq 3\times$ in Brown corpus considered worthwhile for inclusion
- inclusion not always possible: 'better' hardly formalizable without context (should be handled at NL level: e.g. CELT)

Domain Ontologies

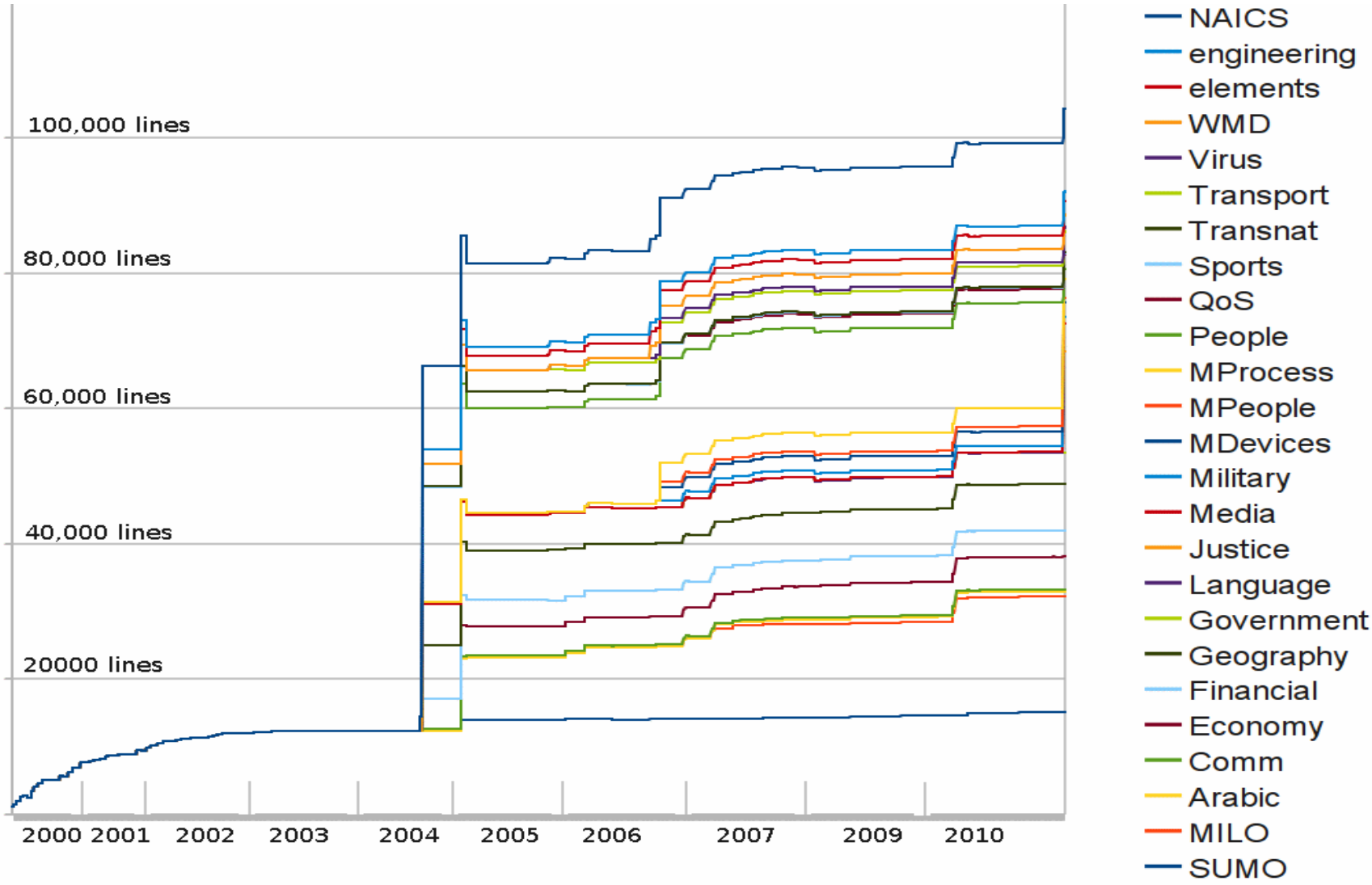
- deliveries for researchers and projects: e.g. Geography Ontology (for government), Media Ontology (for contractor in London)

What goes in SUMO and what in MILO?

- no 'deep' answer
- simply keep SUMO around 1000 terms
- hand-select most specific terms in SUMO and move to MILO

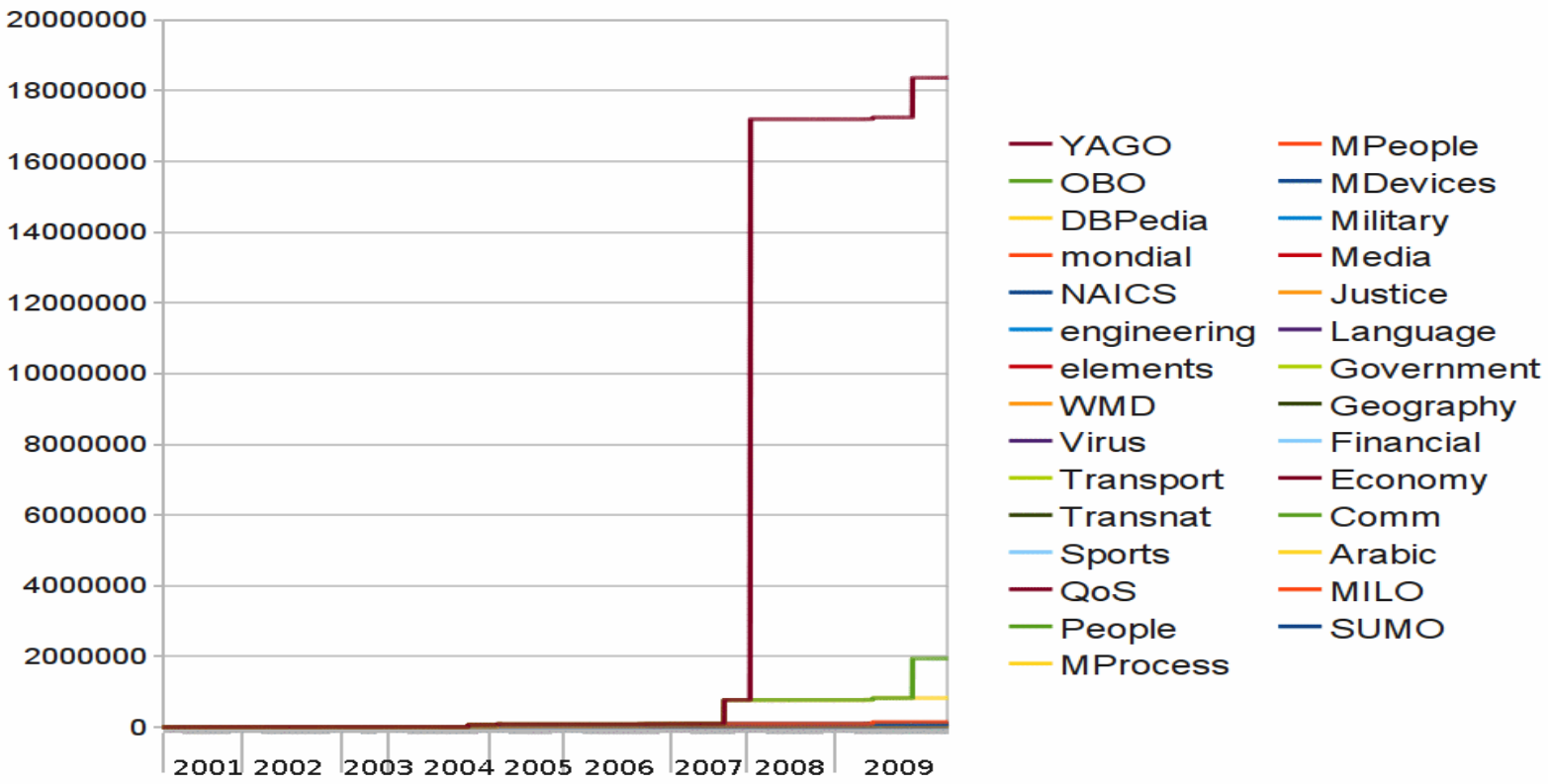
The Growth of SUMO

The Growth of SUMO



The Growth of SUMO

With YAGO, OBO, DBPedia, mondial
(semi-automatic, partial inclusions)



Tool Support (not only) for SUMO Development

Tool Support (not only) for SUMO Development

Sigma Knowledge Engineering Tool

- browsing and displaying tool
- supports inspection and debugging
- open source: `sigmakee.sourceforge.net`

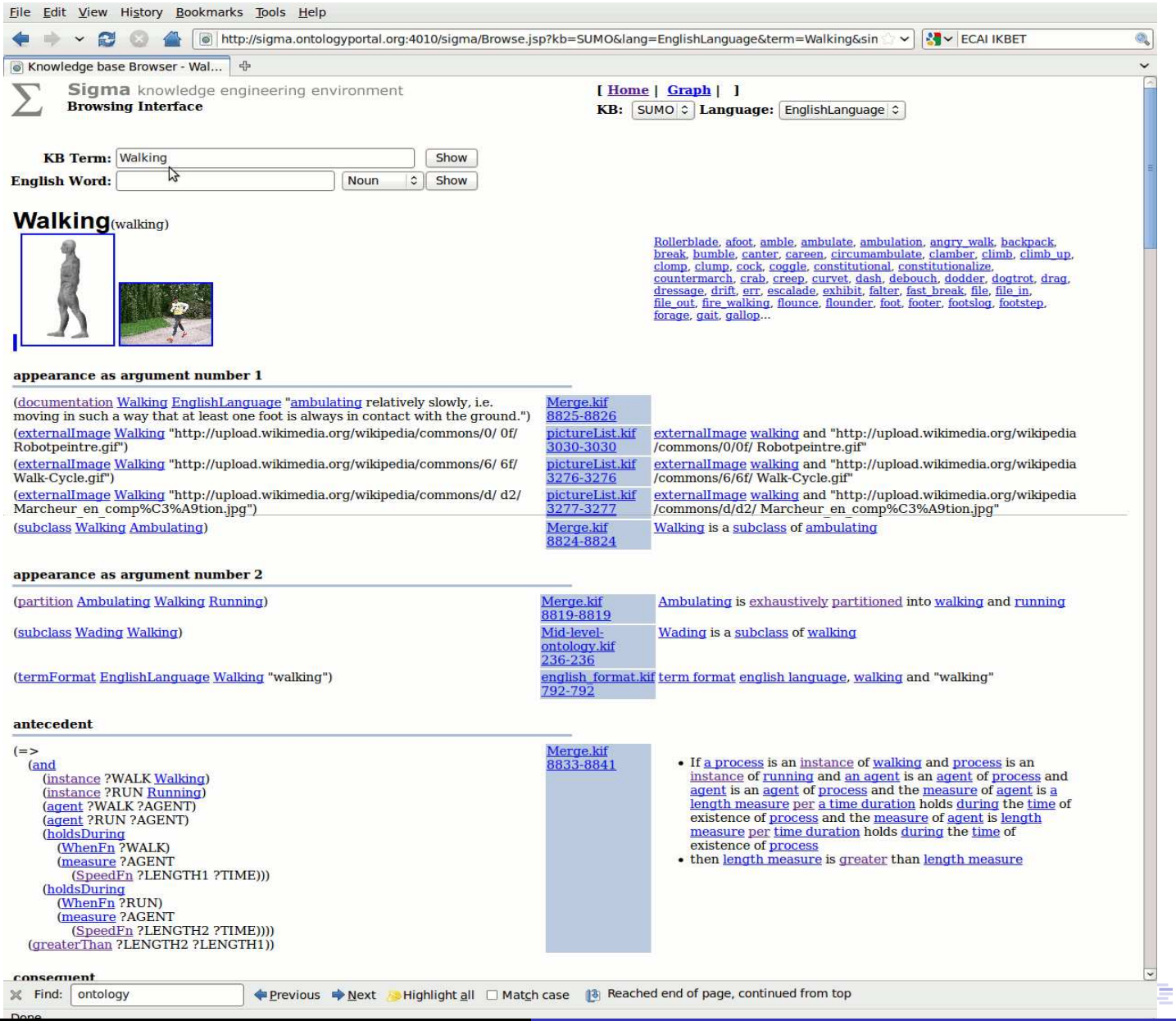
FOL & HOL Theorem Provers — Integrated with Sigma

- KIFVampire (FOL), SystemOnTPTP (FOL), LEO-II (HOL)
- support for ontology debugging and question answering

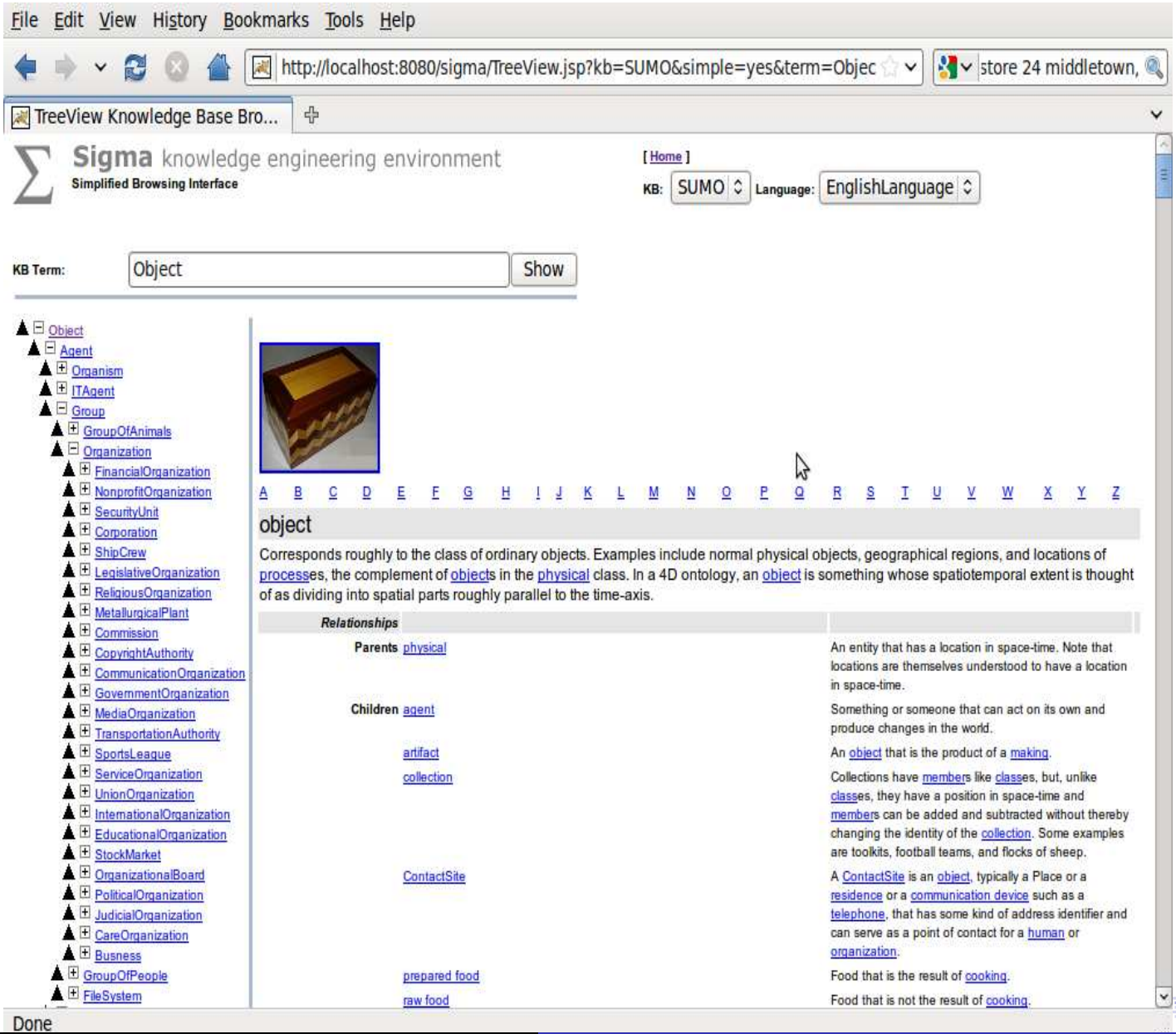
CELT

- Deep NL processing with respect to own ontology
  (cf. LogAnswer which does not have its own ontology)

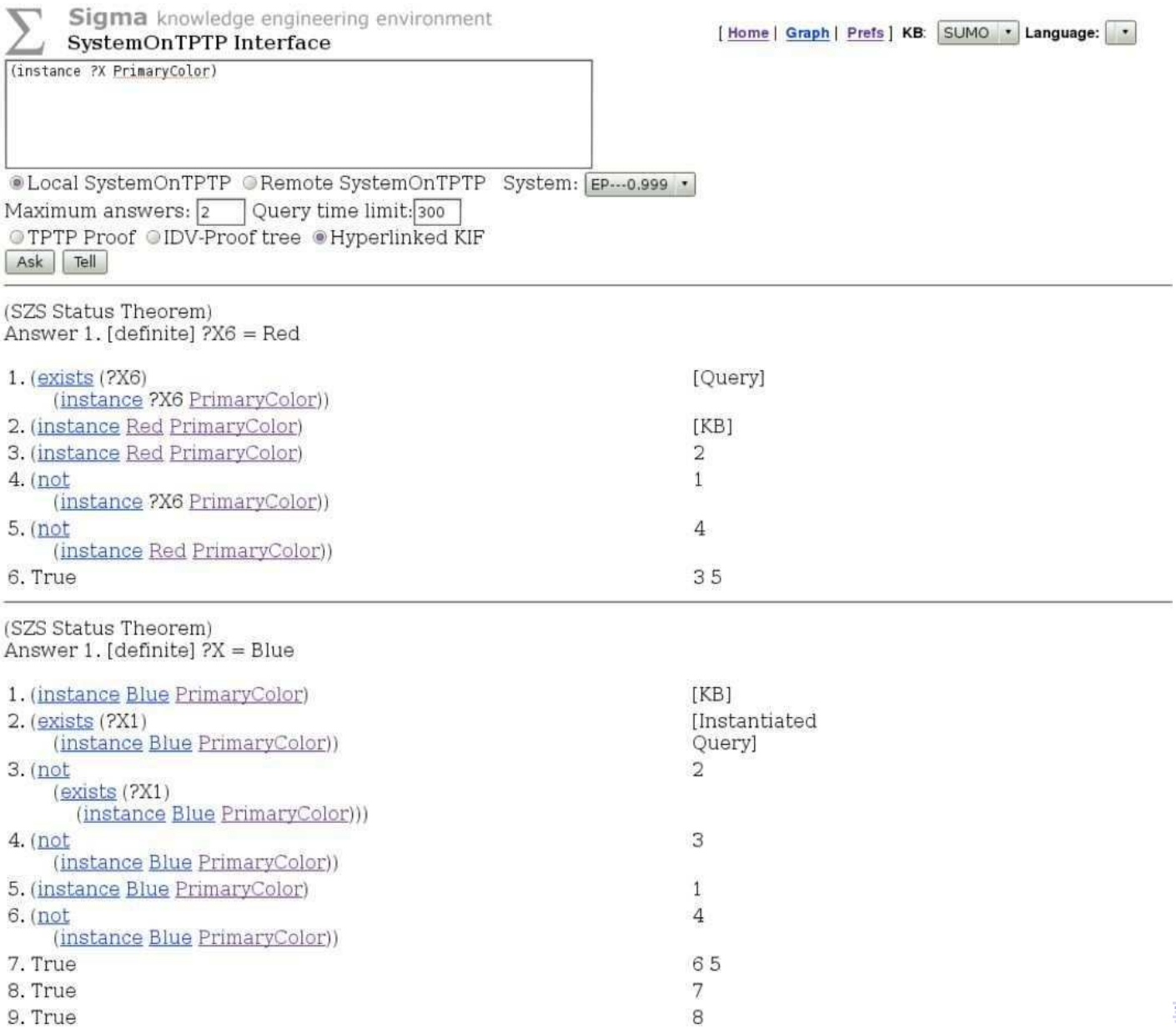
Tool Support (not only) for SUMO Development



Tool Support (not only) for SUMO Development



Tool Support (not only) for SUMO Development



Sigma knowledge engineering environment
SystemOnTPTP Interface

[Home | Graph | Prefs] KB: SUMO Language:

```
(instance ?X PrimaryColor)
```

Local SystemOnTPTP Remote SystemOnTPTP System: EP---0.999

Maximum answers: 2 Query time limit: 300

TPTP Proof IDV-Proof tree Hyperlinked KIF

Ask Tell

(SZS Status Theorem)
Answer 1. [definite] ?X6 = Red

1. (exists (?X6) (instance ?X6 PrimaryColor))	[Query]
2. (instance Red PrimaryColor)	[KB]
3. (instance Red PrimaryColor)	2
4. (not (instance ?X6 PrimaryColor))	1
5. (not (instance Red PrimaryColor))	4
6. True	3 5

(SZS Status Theorem)
Answer 1. [definite] ?X = Blue

1. (instance Blue PrimaryColor)	[KB]
2. (exists (?X1) (instance Blue PrimaryColor))	[Instantiated Query]
3. (not (exists (?X1) (instance Blue PrimaryColor)))	2
4. (not (instance Blue PrimaryColor))	3
5. (instance Blue PrimaryColor)	1
6. (not (instance Blue PrimaryColor))	4
7. True	6 5
8. True	7
9. True	8

Insights, Challenges, Future Work

Insights, Challenges, Future Work

Overall

- biggest surprise so far: no big surprise so far
  (although problem expectation was high for some contractors)
- local problems are continuously being detected and revised
- periodic minor revisions as required/appropriate
- no *fundamental* issues detected so far (which does not mean there are no fundamentally different theories about the world)
- interesting recent problem: semantics of embedded formulas and modal operators $\longrightarrow$ solution proposed in next talk
- interesting and demanding nature of field
- SUMO interesting and relevant to many communities: Computer Science, Linguistics, Philosophy (hard though for students to get work accepted as MSc, PhD)
- example use in Linguistics: checking analogies in NL

Insights, Challenges, Future Work

Overall

- biggest surprise so far: no big surprise so far
  (although problem expectation was high for some contractors)
- local problems are continuously being detected and revised
- periodic minor revisions as required/appropriate
- no *fundamental* issues detected so far (which does not mean there are no fundamentally different theories about the world)
- interesting recent problem: semantics of embedded formulas and modal operators $\longrightarrow$ solution proposed in next talk
- interesting and demanding nature of field
- SUMO interesting and relevant to many communities: Computer Science, Linguistics, Philosophy (hard though for students to get work accepted as MSc, PhD)
- example use in Linguistics: checking analogies in NL

Insights, Challenges, Future Work

Overall

- biggest surprise so far: no big surprise so far
  (although problem expectation was high for some contractors)
- local problems are continuously being detected and revised
- periodic minor revisions as required/appropriate
- no *fundamental* issues detected so far (which does not mean there are no fundamentally different theories about the world)
- interesting recent problem: semantics of embedded formulas and modal operators $\longrightarrow$ solution proposed in next talk
- interesting and demanding nature of field
- SUMO interesting and relevant to many communities: Computer Science, Linguistics, Philosophy (hard though for students to get work accepted as MSc, PhD)
- example use in Linguistics: checking analogies in NL

Insights, Challenges, Future Work

Overall

- biggest surprise so far: no big surprise so far
  (although problem expectation was high for some contractors)
- local problems are continuously being detected and revised
- periodic minor revisions as required/appropriate
- no *fundamental* issues detected so far (which does not mean there are no fundamentally different theories about the world)
- interesting recent problem: semantics of embedded formulas and modal operators $\longrightarrow$ solution proposed in next talk
- interesting and demanding nature of field
- SUMO interesting and relevant to many communities: Computer Science, Linguistics, Philosophy (hard though for students to get work accepted as MSc, PhD)
- example use in Linguistics: checking analogies in NL

Insights, Challenges, Future Work

Overall

- biggest surprise so far: no big surprise so far
  (although problem expectation was high for some contractors)
- local problems are continuously being detected and revised
- periodic minor revisions as required/appropriate
- no *fundamental* issues detected so far (which does not mean there are no fundamentally different theories about the world)
- interesting recent problem: semantics of embedded formulas and modal operators $\longrightarrow$ solution proposed in next talk
- interesting and demanding nature of field
- SUMO interesting and relevant to many communities: Computer Science, Linguistics, Philosophy (hard though for students to get work accepted as MSc, PhD)
- example use in Linguistics: checking analogies in NL

Insights, Challenges, Future Work

Open Source

- SUMO is largest open source formal ontology
- motivation: make SUMO ubiquous, hope for real impact
- disappointment: not the big growth as originally anticipated
- some contractors do not want their additions to become open source $\longrightarrow$ no need to publish new domain ontologies
- concrete license: BSD like license for SUMO, GPL for MILO (creative commons license would be choice today)

Insights, Challenges, Future Work

Open Source

- SUMO is largest open source formal ontology
- motivation: make SUMO ubiquous, hope for real impact
- disappointment: not the big growth as originally anticipated
- some contractors do not want their additions to become open source $\longrightarrow$ no need to publish new domain ontologies
- concrete license: BSD like license for SUMO, GPL for MILO (creative commons license would be choice today)

Ongoing and Future Work

- better archeology (cvs logs available for purpose): detailed analysis/inspection of modifications to
  - comments
  - hierarchy
  - axioms
  - ...

What can be learned?

- proper treatment of modal operators $\longrightarrow$ possible world semantics for SUMO?
- appropriate treatment of (other) higher-order aspects
- modeling of interesting problem scenarios (e.g. Smullyan's puzzles) and testing of SUMO and Sigma for them
- more domain ontologies
- use of CELT & SUMO in NL applications

recruitment of students

Ongoing and Future Work

- better archeology (cvs logs available for purpose): detailed analysis/inspection of modifications to
  - comments
  - hierarchy
  - axioms
  - ...

  What can be learned?
- proper treatment of modal operators $\longrightarrow$ possible world semantics for SUMO?
- appropriate treatment of (other) higher-order aspects
- modeling of interesting problem scenarios (e.g. Smullyan's puzzles) and testing of SUMO and Sigma for them
- more domain ontologies
- use of CELT & SUMO in NL applications

recruitment of students

Ongoing and Future Work

- better archeology (cvs logs available for purpose): detailed analysis/inspection of modifications to
  - comments
  - hierarchy
  - axioms
  - ...

  What can be learned?
- proper treatment of modal operators $\longrightarrow$ possible world semantics for SUMO?
- appropriate treatment of (other) higher-order aspects
- modeling of interesting problem scenarios (e.g. Smullyan's puzzles) and testing of SUMO and Sigma for them
- more domain ontologies
- use of CELT & SUMO in NL applications

recruitment of students

Insights, Challenges, Future Work

Ongoing and Future Work

- better archeology (cvs logs available for purpose): detailed analysis/inspection of modifications to
  - comments
  - hierarchy
  - axioms
  - ...

  What can be learned?
- proper treatment of modal operators $\longrightarrow$ possible world semantics for SUMO?
- appropriate treatment of (other) higher-order aspects
- modeling of interesting problem scenarios (e.g. Smullyan's puzzles) and testing of SUMO and Sigma for them
- more domain ontologies
- use of CELT & SUMO in NL applications

recruitment of students

Ongoing and Future Work

- better archeology (cvs logs available for purpose): detailed analysis/inspection of modifications to
  - comments
  - hierarchy
  - axioms
  - ...

  What can be learned?
- proper treatment of modal operators $\longrightarrow$ possible world semantics for SUMO?
- appropriate treatment of (other) higher-order aspects
- modeling of interesting problem scenarios (e.g. Smullyan's puzzles) and testing of SUMO and Sigma for them
- more domain ontologies
- use of CELT & SUMO in NL applications

Ongoing and Future Work

recruitment of students

- better archeology (cvs logs available for purpose): detailed analysis/inspection of modifications to
  - comments
  - hierarchy
  - axioms
  - ...

  What can be learned?
- proper treatment of modal operators $\longrightarrow$ possible world semantics for SUMO?
- appropriate treatment of (other) higher-order aspects
- modeling of interesting problem scenarios (e.g. Smullyan's puzzles) and testing of SUMO and Sigma for them
- more domain ontologies
- use of CELT & SUMO in NL applications

Insights, Challenges, Future Work

Challenge

expressive, general purpose knowledge representation

robust and effective integration

heterogeneous reasoning systems (general purpose & specialist)

Thank You!