

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

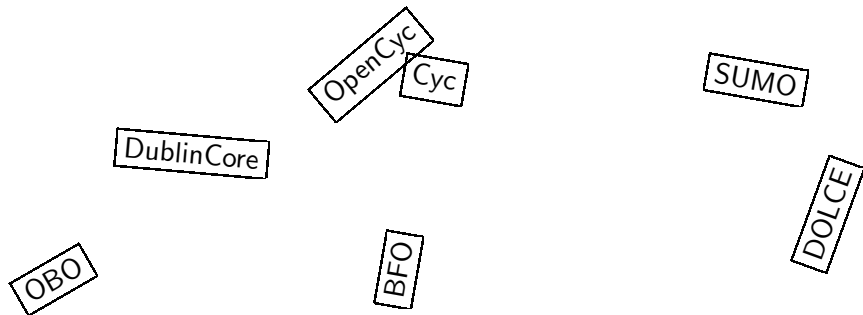
Adam Pease and Christoph Benz Müller*

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

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)

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)

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

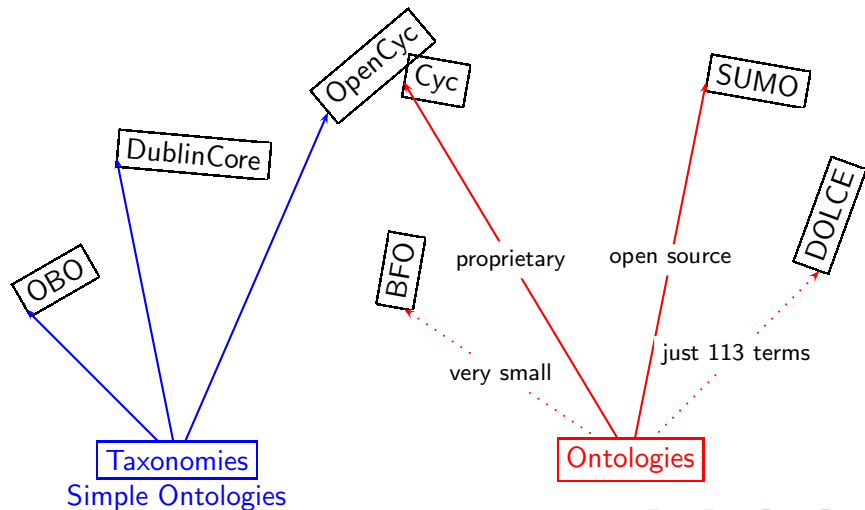
Characteristics

- ▶ 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?

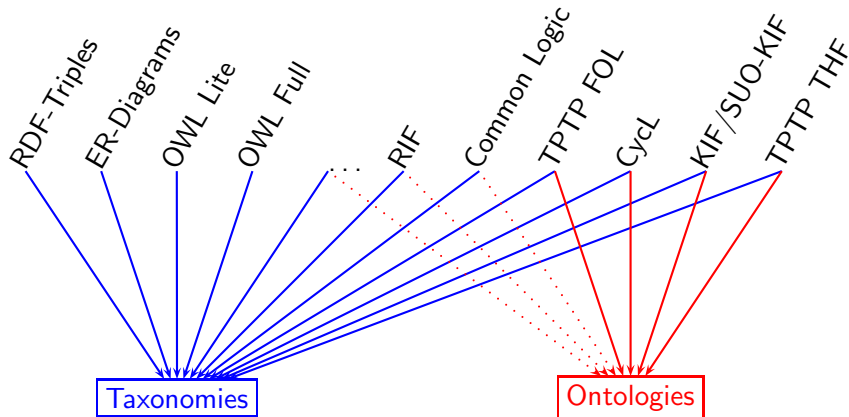
- ▶ ontology development \approx
unrestricted, large-scale software development

Ontology: An (Over-)Stretched Notion?



Ontology: An (Over-)Stretched Notion?

sufficiently expressive



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)
```

...

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

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

- ▶ 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

- ▶ 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

- ▶ 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

- ▶ 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

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 — '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)

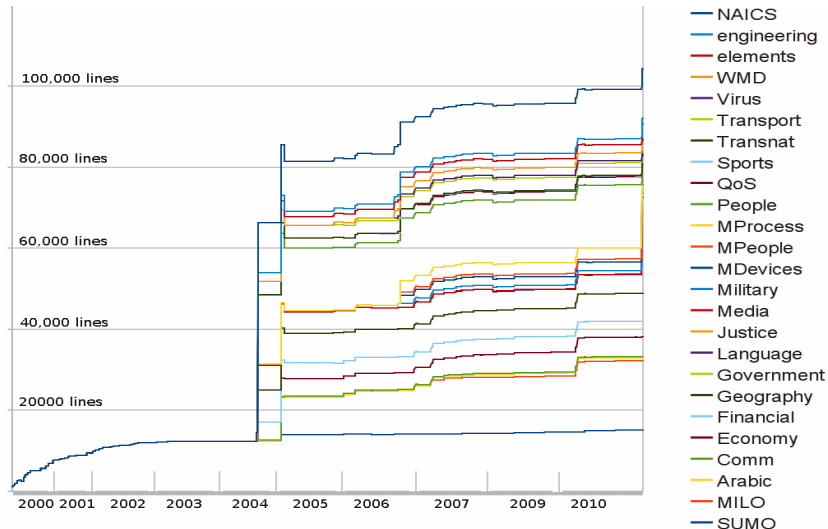
What goes in SUMO and what in MILO?

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



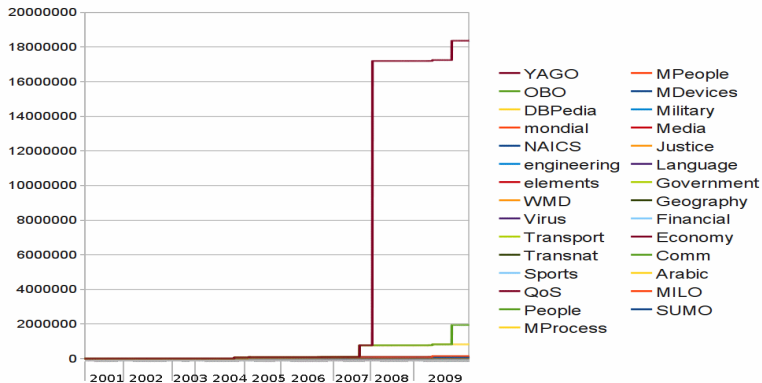
The Growth of SUMO

The Growth of SUMO



The Growth of SUMO

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





Tool Support (not only) for **SUMO** Development

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

The screenshot shows the Sigma knowledge engineering environment interface. At the top, there is a menu bar with 'File', 'Edit', 'View', 'History', 'Bookmarks', 'Tools', and 'Help'. Below the menu bar is a browser address bar showing the URL: <http://sigma.ontologyportal.org:4010/sigma/Browse.jsp?kb=SUMO&lang=EnglishLanguage&term=Walking&in>. The main content area is titled 'Sigma knowledge engineering environment' and 'Browsing Interface'. It features a search bar with 'KB Term: Walking' and 'English Word: Noun'. Below the search bar, the entry for 'Walking (walking)' is displayed. The entry includes two images: a silhouette of a person walking and a photograph of a person walking in a park. To the right of the images is a list of related terms: Rollerblade, about, amble, ambulate, ambulation, angry walk, backpack, break, bump, center, career, circumambulate, clamber, climb, climb up, climb, clump, creak, creak, constitutional, constitutionalis, counter-march, crab, creep, curvet, dash, dehoosh, dodder, dog trot, drag, message, drift, err, escalate, exhibit, falter, fast break, file, file in, file out, fire walking, flourish, founder, foot, footer, footstep, footstep, forage, gait, gallop... Below the images and related terms, the entry is divided into sections: 'appearance as argument number 1', 'appearance as argument number 2', and 'antecedent'. Each section contains a list of external image links and a table of internal links. The 'antecedent' section contains a list of logical expressions and a table of internal links. At the bottom of the interface, there is a search bar with 'Find: ontology' and a set of navigation buttons: Previous, Next, Highlight all, Match case, and Reached end of page, continued from top.

KB Term: Show

English Word: Show

Walking (walking)



Rollerblade, about, amble, ambulate, ambulation, angry walk, backpack, break, bump, center, career, circumambulate, clamber, climb, climb up, climb, clump, creak, creak, constitutional, constitutionalis, counter-march, crab, creep, curvet, dash, dehoosh, dodder, dog trot, drag, message, drift, err, escalate, exhibit, falter, fast break, file, file in, file out, fire walking, flourish, founder, foot, footer, footstep, footstep, forage, gait, gallop...

appearance as argument number 1

(documentation Walking EnglishLanguage "ambulating relatively slowly, i.e. moving in such a way that at least one foot is always in contact with the ground.")

externalImage Walking "http://upload.wikimedia.org/wikipedia/commons/0/0f/Robotpeintre.gif"	Merge.kif 8525-8526	externalImage walking and "http://upload.wikimedia.org/wikipedia/commons/0/0f/ Robotpeintre.gif"
externalImage Walking "http://upload.wikimedia.org/wikipedia/commons/6/6f/ Walk-Cycle.gif"	pictureList.kif 3030-3030	externalImage walking and "http://upload.wikimedia.org/wikipedia/commons/6/6f/ Walk-Cycle.gif"
externalImage Walking "http://upload.wikimedia.org/wikipedia/commons/d/d2/ Marcheur_en_comp%C3%A9tion.jpg"	pictureList.kif 3276-3276	externalImage walking and "http://upload.wikimedia.org/wikipedia/commons/d/d2/ Marcheur_en_comp%C3%A9tion.jpg"
(subclass Walking Ambulating)	pictureList.kif 3277-3277	externalImage walking and "http://upload.wikimedia.org/wikipedia/commons/d/d2/ Marcheur_en_comp%C3%A9tion.jpg"
	Merge.kif 8824-8824	Walking is a subclass of ambulating

appearance as argument number 2

(partition Ambulating Walking Running)

(subclass Wading Walking)	Merge.kif 8819-8819	Ambulating is exhaustively partitioned into walking and running
(termFormat EnglishLanguage Walking "walking")	Mid-level-ontology.kif 236-236	Wading is a subclass of walking
	english_format.kif 792-792	term format english language: walking and "walking"

antecedent

(=>

- [land](#)
- [\(instance ?WALK Walking\)](#)
- [\(instance ?RUN Running\)](#)
- [\(agent ?WALK ?AGENT\)](#)
- [\(agent ?RUN ?AGENT\)](#)
- [\(holdsDuring](#)
- [\(WhenEn ?WALK](#)
- [\(measure ?AGENT](#)
- [\(SpeedEn ?LENGTH1 ?TIME\)\)\)](#)
- [\(holdsDuring](#)
- [\(WhenEn ?RUN](#)
- [\(measure ?AGENT](#)
- [\(SpeedEn ?LENGTH2 ?TIME\)\)\)](#)
- [\(greaterThan ?LENGTH2 ?LENGTH1\)\)](#)

Merge.kif 8833-8841	<ul style="list-style-type: none">If a process is an instance of walking and process is an instance of running and an agent is an agent of process and agent is an agent of process and the measure of agent is a length measure per a time duration holds during the time of existence of process and the measure of agent is length measure per time duration holds during the time of existence of processThen length measure is greater than length measure
-------------------------------------	--

Find: ontology Previous Next Highlight all Match case Reached end of page, continued from top

Tool Support (not only) for SUMO Development


The screenshot shows the Sigma knowledge engineering environment. The browser address bar indicates the URL: `http://localhost:8080/sigma/TreeView.jsp?kb=SUMO&simple=yes&term=Object`. The page title is "Sigma knowledge engineering environment" and the subtitle is "Simplified Browsing Interface". The KB is set to "SUMO" and the Language is "EnglishLanguage".

The "KB Term" is "Object". A tree view on the left shows a hierarchy of classes, including "Object", "Agent", "Organization", "Group", "GroupOfAnimals", "FinancialOrganization", "NonprofitOrganization", "SecurityUnit", "Corporation", "ShipCrew", "LegislativeOrganization", "ReligiousOrganization", "MetalurgicalPlant", "Commission", "CopyrightAuthority", "CommunicationOrganization", "GovernmentOrganization", "MediaOrganization", "TransportationAuthority", "SportsLeague", "ServiceOrganization", "UnionOrganization", "InternationalOrganization", "EducationalOrganization", "StockMarket", "OrganizationalBoard", "PoliticalOrganization", "JudicialOrganization", "CareOrganization", "Business", "GroupOfPeople", and "FileSystem".

The main content area displays the "object" class. It includes a small image of a cake and a table of relationships. The table has columns for "Relationships" and "Description".

Relationships	Description
Parents physical	An entity that has a location in space-time. Note that locations are themselves understood to have a location in space-time.
Children agent	Something or someone that can act on its own and produce changes in the world.
artifact	An object that is the product of a making .
collection	Collections have members like classes , but, unlike classes , they have a position in space-time and members can be added and subtracted without thereby changing the identity of the collection . Some examples are toolkits, football teams, and flocks of sheep.
ContactSite	A ContactSite is an object , typically a Place or a residence or a communication device , such as a telephone , that has some kind of address identifier and can serve as a point of contact for a human or organization .
prepared food	Food that is the result of cooking .
raw food	Food that is not the result of cooking .

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: Query time limit:


TPTP Proof IDV-Proof tree Hyperlinked KIF

(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

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 → **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

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 → **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

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 → **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

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 → **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

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 → **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

Open Source

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

Open Source

- ▶ SUMO is largest open source formal ontology
- ▶ motivation: make SUMO ubiquitous, hope for real impact
- ▶ disappointment: not the big growth as originally anticipated
- ▶ some contractors do not want their additions to become open source → 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

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 → 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 → 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 → 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 → 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 → 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 → 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

Challenge

expressive, general purpose knowledge representation



robust and effective integration

heterogeneous reasoning systems (general purpose & specialist)

Thank You!