The Modular Structure of an Ontology: Atomic Decomposition

Chiara Del Vescovo[1] Bijan Parsia[1]
Uli Sattler[1] *Thomas Schneider*[2]

[1]School of Computer Science, University of Manchester, UK

[2]Dept. of Computer Science, University of Bremen, Germany

LogInf, 4 November 2011

Ontologies

. . . are finite theories in a (description) logic, e.g.

$$\begin{aligned} \mathsf{Part} &\equiv \exists \mathsf{strict\_part\_of}.\mathsf{Whole} \\ \mathsf{strict\_part\_of} &\sqsubseteq \mathsf{part\_of} \\ \mathsf{Hand} &\sqsubseteq {=}5\,\mathsf{part\_of}.\mathsf{Finger} \end{aligned}$$

OWL (W3C standard, FOL fragment): expressive DL, featuring

- unary and binary predicates
- constructors for unary predicates

  Booleans, $\exists$, $\forall$, counting quantifiers, nominals
- constructors for binary predicates

  inverse, composition
- axioms: inclusions/equivalences of predicates
- global constraints to ensure decidability

Modules

A module $\mathcal{M}(\Sigma, \mathcal{O}) \subseteq \mathcal{O}$ encapsulates knowledge
w.r.t. a signature Σ if $\mathcal{M} \equiv^{c}_{\Sigma} \mathcal{O}$

i.e., for all $C \sqsubseteq D$ with $\mathrm{sig}(C \sqsubseteq D) \subseteq \Sigma$:
$\mathcal{O} \models C \sqsubseteq D$ iff $\mathcal{M}(\Sigma, \mathcal{O}) \models C \sqsubseteq D$

Example: $\mathcal{O} = \texttt{Mereology.owl}$, $\Sigma = \{\mathsf{Part}, \mathsf{Whole}\}$, $\mathcal{M}(\Sigma, \mathcal{O}) =$

$\mathcal{O}$

$\mathcal{M}$

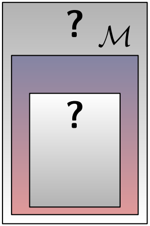
$$\begin{array}{rlrl}
\mathsf{Part} \equiv & \exists \mathsf{strict_part_of}.\mathsf{Whole} & \mathsf{strict_part_of} \equiv & \mathsf{direct_part_of} \\
\mathsf{Part} \sqsubseteq & \forall \mathsf{strict_part_of}.\mathsf{Whole} & \mathsf{direct_part_of} \equiv & \mathsf{direct_part}^{-1} \\
\mathsf{Whole} \equiv & \exists \mathsf{strict_part}.\mathsf{Part} & \mathsf{strict_part} \equiv & \mathsf{direct_part} \\
\mathsf{Whole} \sqsubseteq & \forall \mathsf{strict_part}.\mathsf{Part} & \mathsf{strict_part_of} \equiv & \mathsf{strict_part}^{-1}
\end{array}$$

Modular *structure*

Modules are great: if you know your (seed) signature . . .
and for "module local" tasks such as reuse

Single module extraction does *not* help if you

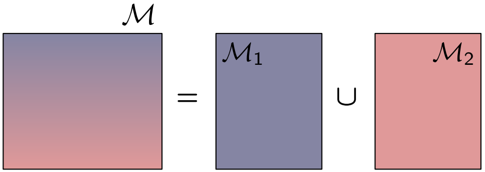
- do *not* know the *right* seed signature
- want to understand *other* modules
- want to understand *axiom dependency structure*



To analyse the modular structure of the ontology:

- significant modules
- significant relations between modules
- . . . which reveals logical dependency between axioms

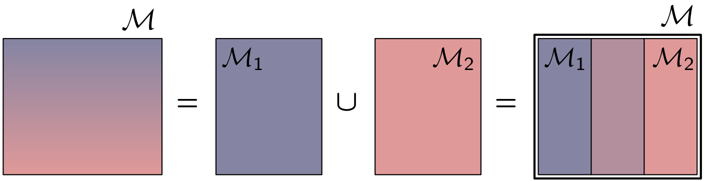
Are all modules significant?



To understand $\mathcal{M}$, one must understand

- the dependency structure of $\mathcal{M}_1$
- the dependency structure of $\mathcal{M}_2$
- nothing else: $\mathcal{M}_1$ and $\mathcal{M}_2$ have no further dependencies

Are all modules significant?



To understand $\mathcal{M}$, one must understand

- the dependency structure of $\mathcal{M}_1$
- the dependency structure of $\mathcal{M}_2$
- nothing else: $\mathcal{M}_1$ and $\mathcal{M}_2$ have no further dependencies

$\rightsquigarrow$ $\mathcal{M}$ is not significant: it is a fake module

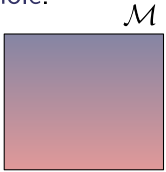
- $\mathcal{M}_1$ and $\mathcal{M}_2$ may be "significant"
- Knowing that $\mathcal{M}$ is "only" a union is important

Are all modules significant?

Consider a module $\mathcal{M}$ that is not fake.

To understand $\mathcal{M}$, one has to understand $\mathcal{M}$ as a whole.

- all axioms in $\mathcal{M}$ logically interact
- in different ways – but interact



"Not fake" implies significant: genuine

Ratio of fake to genuine

Given a set of genuine modules,

- Unions lead to fake modules

$\rightsquigarrow$ The space of fake modules is exponential

- But not *every* union of genuine modules is a module

The number of *all* modules can and does grow exponentially in $|\mathcal{O}|$
[D.,P.,S.,S., KR 2010 & WoMO 2010]

Question 1

Is module growth primarily due to trivial combinations?
I.e., are most modules fake?

Yes!

Theorem 1

Each genuine module is the smallest module that contains α, for some axiom $\alpha \in \mathcal{O}$.

$\leadsto$ The family of genuine modules is linear in $|\mathcal{O}|$.

Most modules are fake!

Proof exploits properties of modules

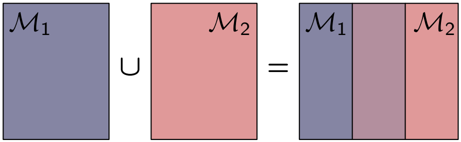
- uniqueness, monotonicity, self-containedness, ...
- which are satisfied by all locality-based modules

Relations between modules

Genuine modules may overlap.

This exposes significant logical dependency between axioms:

Axioms in $\mathcal{M}_1 \setminus \mathcal{M}_2$ depend on axioms in $\mathcal{M}_1 \cap \mathcal{M}_2$

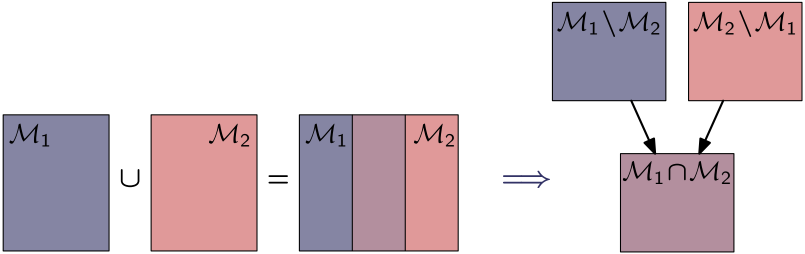


Relations between modules

Genuine modules may overlap.

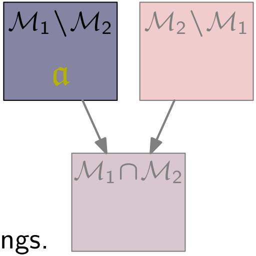
This exposes significant logical dependency between axioms:

Axioms in $\mathcal{M}_1 \setminus \mathcal{M}_2$ depend on axioms in $\mathcal{M}_1 \cap \mathcal{M}_2$



Atoms

An atom is a maximal set $\mathfrak{a} \subseteq \mathcal{O}$ such that, for every module $\mathcal{M}$, either $\mathfrak{a} \subseteq \mathcal{M}$ or $\mathfrak{a} \cap \mathcal{M} = \emptyset$.



- The smallest module for an axiom α contains the whole atom to which α belongs.
- Axioms in an atom are logically interdependent.
- Any two atoms are disjoint.

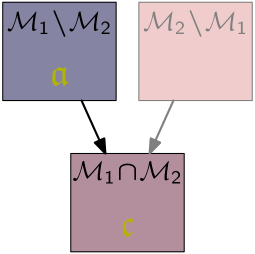
$\rightsquigarrow$ The family of atoms is a partition of the ontology.

- Each module is a disjoint union of atoms.

Proposition

There is a 1-1 correspondence btwn genuine modules and atoms.

Atomic Decomposition



Dependence between atoms:

- $\mathfrak{a} \succeq \mathfrak{c}$ if, for each $\mathcal{M}$: $\mathfrak{a} \subseteq \mathcal{M}$ implies $\mathfrak{c} \subseteq \mathcal{M}$
- Axioms in $\mathfrak{a}$ logically depend on axioms in $\mathfrak{c}$

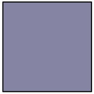
Theorem 2

The relation $\succeq$ is reflexive, antisymmetric, and transitive.

$\rightsquigarrow$ A Hasse diagram exposes 2 logical dependencies: amongst axioms in atoms, between atoms

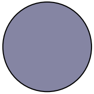
Mereology ontology

42 axioms
1952 modules



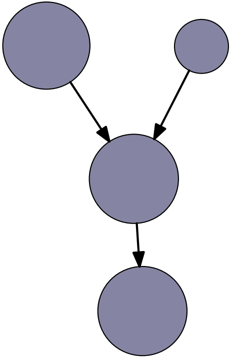
Mereology ontology

42 axioms
1952 modules

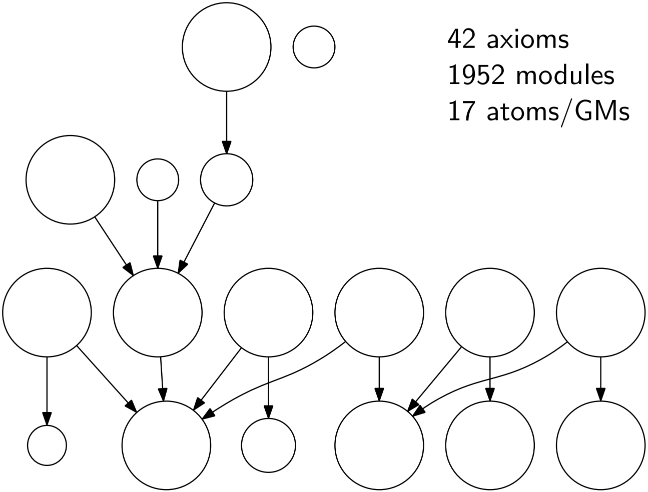


Mereology ontology

42 axioms
1952 modules



Mereology ontology



42 axioms
1952 modules
17 atoms/GMs

Feasibility

Question 2

Can we

- compute all genuine modules
- compute all atoms
- compute their dependencies

without computing all modules?!

Yes!

Remember:

Theorem 1

Each genuine module is the smallest module that contains α, for some axiom $\alpha \in \mathcal{O}$.

- Extract $\mathcal{M}(\mathsf{sig}(\alpha), \mathcal{O})$
  ($\leqslant$ linearly many module extractions)
- Atomic decomposition induced by the comparison of only the genuine modules
  (quadratic procedure)

In Reality?

We have decomposed 181 OWL ontologies from NCBO BioPortal

Decomposability:

Average	nr. axioms per atom	1.73
”	max. nr. axioms per atom	86
”	nr. axioms per genuine module	66
”	max. nr. axioms per genuine mod.	143

Summary

- The atomic decomposition (AD) is a linear representation of the potentially exponential set of all modules.
- AD can be computed using a linear number of module extractions.
- AD exposes 2 types of logical dependencies between axioms.

Future work

- Dependency between atoms and *sets* of atoms
- Labels for atoms – different labels for different tasks
- Applications
  - Topicality for ontology comprehension
  - Fast module extraction
  - All module count
  - ...

Future work

- Dependency between atoms and *sets* of atoms
- Labels for atoms – different labels for different tasks
- Applications
  - Topicality for ontology comprehension
  - Fast module extraction
  - All module count
  - ...

Thank you.

Decomposability issues

Ontology $\mathcal{O}$ (ID in BioPortal)	$\#\mathcal{O}$	#max Atom	#Eq. axs.	#Disj. axs.
Nanoparticle Ontology (1083)	16,267	6,425	42	6,106
Breast Tissue Cell Lines Ontology (1438)	2,734	2,201	0	7
IMGT Ontology (1491)	1,112	729	38	594
SNP Ontology (1058)	3,481	598	30	210
Amino Acid Ontology (1054)	477	445	8	190
Comparative Data Analysis (1128)	804	434	8	190
Family Health History (1126)	1,091	378	0	1
Neural Electromagnetic Ontologies (1321)	2,286	259	21	0
Computer-based Patient Record Ontology (1059)	1,454	238	18	20
Basic Formal Ontology (1332)	95	89	13	41
Ontology of Medically-related Social Entities (1565)	138	100	17	41
Ontology for General Medical Science (1414)	194	102	17	41
Cancer Research and Mgmt Acgt Master (1130)	5,435	3,796	16	42