The Modular Structure of an Ontology: Atomic Decomposition

Chiara Del Vescovo¹  Bijan Parsia¹
Uli Sattler¹  Thomas Schneider²

¹The University of Manchester, UK
²Universität Bremen, Germany
Ontologies & Modules

- **An ontology** is a finite set of axioms in a (description) logic

- A **module** $M(\Sigma, O) \subseteq O$ encapsulates knowledge w.r.t. a signature $\Sigma$: $M \equiv^c_\Sigma O$

  i.e., for all $C \subseteq D$ with $\text{sig}(C \subseteq D) \subseteq \Sigma$:

  $O \models C \subseteq D$ iff $M(\Sigma, O) \models C \subseteq D$

$M(\{\text{part}\}, \text{Mereology.owl}) = \{\text{Trans: part, part InverseOf: PartOf, Trans: partOf}\}$
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 dependence between axioms
Are all modules significant?

To understand $M$, one must
- understand the dependency structure of $M_1$
- understand the dependency structure of $M_2$
- nothing else: $M_1$ and $M_2$ have no further dependencies

$M$ is **not** significant: it is a **fake** module
- Thus, $M_1$ and $M_2$ may be “significant”
- knowing that $M$ is “only” a union is important
Are all modules significant?

- To understand $M$, one must
  - understand the dependency structure of $M_1$
  - understand the dependency structure of $M_2$
  - nothing else: $M_1$ and $M_2$ have no further dependencies

  - $M$ is not significant: it is a **fake** module
    - Thus, $M_1$ and $M_2$ may be “significant”
    - knowing that $M$ is “only” a union is important
Are all modules significant?

- Consider a module $M$ that is **not fake**
- To understand $M$, one has to understand $M$ as a whole
  - all axioms in $M$ 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,
  - the space of fake modules is large (exponential)
  - but not every union of genuine modules is a module
- The cardinality of the set of all modules can and does grow exponentially in the size of $O$
  - See D., P., S., S., KR 2010 & WoMO 2010
- *Is module growth primarily due to trivial combinations?*
  - are most modules **fake**?
Theorem 1: Each genuine module is the smallest module for some axiom $\alpha \in O$.

- The family of genuine modules is linear in $|O|$
- 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 dependence between axioms:
  ▪ axioms in $M_1 \setminus M_2$ depend on axioms in $M_1 \cap M_2$
Relations between Modules

- Genuine modules may overlap
- This exposes significant logical dependence between axioms:
  - axioms in $M_1 \setminus M_2$ depend on axioms in $M_1 \cap M_2$
Atoms

- \( \hat{A} \subseteq O \) is an **atom** if it is a maximal set s.t., for each module \( M \), either \( \hat{A} \subseteq M \) or \( \hat{A} \cap M = \emptyset \).
- The smallest module for an axiom \( \alpha \) contains the whole atom to which \( \alpha \) belongs!
- Axioms in an atom are logically interdependent
- Any two atoms are disjoint
- The family of atoms is a partition of the ontology
  - Only linearly many atoms
- Each GM is a disjoint union of atoms

**Proposition:** There is a 1-1 correspondence between genuine modules and atoms.
Atomic Decomposition

- Dependence between atoms:
  - $\hat{A} \succeq \hat{C}$ if, for each $M$: $\hat{A} \subseteq M$ implies $\hat{C} \subseteq M$
  - Axioms in $\hat{A}$ logically depend on axioms in $\hat{C}$

Theorem 2: The relation $\succeq$ is reflexive, antisymmetric, and transitive.

- 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
Can we compute all genuine modules?
- and all atoms
- with their dependencies?
- ...without computing all modules?!
Yes!

- Remember:

Theorem 1: Each genuine module is the smallest module for some axiom \( \alpha \in O \).

- extract \( M(\text{sig}(\alpha), O) \)
  - \( \leq \) linearly many module extractions
- AD induced by the comparison of GMs
  - quadratic procedure
In Reality?

- We have decomposed 181 OWL ontologies from NCBO BioPortal
- Decomposability: average
  - nr. axioms/atom: 1.73
  - max nr. axioms/atom: 86
  - nr. axioms/GM: 66
  - max nr. axioms/GM: 143
Future Work

- More on dependency of axioms
  - between atoms and \textit{sets} of atoms
- Labels for atoms
  - different labels for different tasks
- Applications
  - All Module Count
  - Fast Module Extraction
  - Topicality for Ontology Comprehension: see ICCS 2011
- ...
Thank you! – Questions?
Decomposability Issues

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