What is my ontology about? How many modules? Relevant modules

The Modular Structure of an Ontology: an Empirical Study

Chiara Del Vescovo  Bijan Parsia  Uli Sattler  Thomas Schneider

School of Computer Science, University of Manchester, UK

DL, 7 May 2010  WoMo, 11 May 2010
Three questions

What is my ontology about?

How many modules does my ontology have?

How do we identify relevant modules?
What is my ontology about? How many modules? Relevant modules

And now . . .

What is my ontology about?

How many modules does my ontology have?

How do we identify relevant modules?
What is my ontology about?

We can’t inspect all its axioms.

1,000,000
What is my ontology about?

We can inspect its modular structure, obtained a posteriori.
We bet Robert Stevens

- Ontology about periodic table of the chemical elements
- Logical structure $\approx$ intended modelling?
  - What is its modular structure?
  - What are its main parts?
What is my ontology about? How many modules? Relevant modules

We bet Robert Stevens

- Ontology about periodic table of the chemical elements
- Logical structure $\approx$ intended modelling?
  - What is its modular structure?
  - What are its main parts?
- Challenge: *automatic* partition into meaningful modules
Modular structure with existing tools

Partition of **Koala** via E-connections in Swoop

- **Animal**
  - **Gender**
  - **Degree**
  - **Habitat**

- Red circle: importing part
- Blue circle: imported but non-importing part
- Green circle: isolated part

→ “imports vocabulary from”
Partition for ontology **Periodic**

- importing part
- imported but non-importing part
- isolated part

→ "imports vocabulary from"
Locality-based modules (LBMs)

Module extraction service

**Input:** ontology $\mathcal{O}$; set $\Sigma$ of terms from $\mathcal{O}$

**Output:** subset $\mathcal{M} = \text{mod}(\Sigma, \mathcal{O})$ of $\mathcal{O}$

**Guarantee:** for all axioms $\alpha$ with $\text{sig}(\alpha) \subseteq \Sigma$:

$$
\mathcal{O} \models \alpha \text{ iff } \mathcal{M} \models \alpha
$$

(Coverage)
Locality-based modules (LBMNs)

Module extraction service

Input: ontology $\mathcal{O}$; set $\Sigma$ of terms from $\mathcal{O}$

Output: subset $\mathcal{M} = \text{mod}(\Sigma, \mathcal{O})$ of $\mathcal{O}$

Guarantee: for all axioms $\alpha$ with $\text{sig}(\alpha) \subseteq \Sigma$:

$$\mathcal{O} \models \alpha \iff \mathcal{M} \models \alpha$$  \hspace{1cm} (Coverage)

Modules providing coverage

- encapsulate knowledge about the topic $\Sigma$
- are important for modular import/reuse:
  “Give me all that $\mathcal{O}$ knows about the topic $\Sigma$”
- are hard to extract if minimality is required
Locality-based modules (LBMs)

Module extraction service

Input: ontology $\mathcal{O}$; set $\Sigma$ of terms from $\mathcal{O}$

Output: subset $\mathcal{M} = \text{mod}(\Sigma, \mathcal{O})$ of $\mathcal{O}$

Guarantee: for all axioms $\alpha$ with $\text{sig}(\alpha) \subseteq \Sigma$:

$$\mathcal{O} \models \alpha \text{ iff } \mathcal{M} \models \alpha$$

(Coverage)

LBMs

- provide coverage and therefore encapsulation
- are not always minimal, but often of reasonable size
- can be efficiently computed
- have important robustness properties
Locality-based modules (LBMs)

Module extraction service

Input: ontology \( \mathcal{O} \); set \( \Sigma \) of terms from \( \mathcal{O} \)

Output: subset \( \mathcal{M} = \text{mod}(\Sigma, \mathcal{O}) \) of \( \mathcal{O} \)

Guarantee: for all axioms \( \alpha \) with \( \text{sig}(\alpha) \subseteq \Sigma \):
\[
\mathcal{O} \models \alpha \iff \mathcal{M} \models \alpha
\]
(Coverage)

General remarks:

- Often \( \text{sig}(\mathcal{M}) \neq \Sigma \)
- Different seed signatures can lead to the same module

\[
\text{sig}(\mathcal{M}) \quad \Sigma_1 \quad \cdots \quad \Sigma_n
\]
Want to extract all (relevant) LBMs in order to:

- obtain a finer-grained analysis
- guide users in choosing the right topic(s)
- draw conclusions on characteristics of an ontology:
Modular structure via LBMs

Want to extract all (relevant) LBMs in order to:

- obtain a finer-grained analysis
- guide users in choosing the right topic(s)
- draw conclusions on characteristics of an ontology:
  - To which extent does \( \mathcal{O} \) cover its topics?
  - How strongly are certain terms connected in \( \mathcal{O} \)?
  - What is the axiomatic richness of \( \mathcal{O} \)?
  - Does \( \mathcal{O} \) have superfluous parts?
  - Agreement between logical and intended intuitive modelling?
And now ...

What is my ontology about?

How many modules does my ontology have?

How do we identify relevant modules?
Obvious lower and upper bounds

- Ontologies of size $n$ can have between 1 and $2^n$ modules.
- (We’re working on tighter bounds.)
Obvious lower and upper bounds

- Ontologies of size $n$ can have between 1 and $2^n$ modules.
- (We’re working on tighter bounds.)
- Do real-life ontologies fall into the worst case?
An algorithm that extracts all modules

Results when applied to two small ontologies:

<table>
<thead>
<tr>
<th>Ontology</th>
<th>#Ax</th>
<th>#Terms</th>
<th>#mods</th>
<th>Theor. Max.</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koala</td>
<td>42</td>
<td>25</td>
<td>3660</td>
<td>33 554 432</td>
<td>9s</td>
</tr>
<tr>
<td>Mereology</td>
<td>44</td>
<td>25</td>
<td>1952</td>
<td>33 554 432</td>
<td>3min</td>
</tr>
</tbody>
</table>
An algorithm that extracts all modules

Results when applied to two small ontologies:

<table>
<thead>
<tr>
<th>Ontology</th>
<th>#Ax</th>
<th>#Terms</th>
<th>#mods</th>
<th>Theor. Max.</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koala</td>
<td>42</td>
<td>25</td>
<td>3660</td>
<td>33 554 432</td>
<td>9s</td>
</tr>
<tr>
<td>Mereology</td>
<td>44</td>
<td>25</td>
<td>1952</td>
<td>33 554 432</td>
<td>3min</td>
</tr>
</tbody>
</table>

1. Are 3660 and 1952 “exponential” numbers?
2. How to filter for interesting modules?
Modularisation of subontologies

- Modularised randomly generated parts of 8 ontologies
- Example growth of module numbers:

  
<table>
<thead>
<tr>
<th>Subontology size</th>
<th>#modules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koala</td>
<td></td>
</tr>
<tr>
<td>People</td>
<td></td>
</tr>
</tbody>
</table>

Trendline equation: $y = O(1.2^x)$
Confidence: 0.90

Trendline equation: $y = O(1.5^x)$
Confidence: 0.96
Modularisation of subontologies

- Modularised randomly generated parts of 8 ontologies
- Example growth of module numbers:

![Graph showing growth of module numbers](image)

- **Koala**: Trendline equation: \( y = O(1.2^x) \)
  - Confidence: 0.90

- **People**: Trendline equation: \( y = O(1.5^x) \)
  - Confidence: 0.96
  - Exponential!
And now...

What is my ontology about?

How many modules does my ontology have?

How do we identify relevant modules?
Unification of similar modules

- Identify sets $\mathcal{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_k\}$ of modules that differ in only few axioms
- Replace $\mathcal{M}$ with $\bigcup \mathcal{M}$ and $\bigcap \mathcal{M}$
Unification of similar modules

- Identify sets $\mathcal{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_k\}$ of modules that differ in only few axioms
- Replace $\mathcal{M}$ with $\bigcup \mathcal{M}$ and $\bigcap \mathcal{M}$
Unification of similar modules

- Identify sets $\mathcal{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_k\}$ of modules that differ in only few axioms
- Replace $\mathcal{M}$ with $\bigcup \mathcal{M}$ and $\bigcap \mathcal{M}$
What is my ontology about?  
How many modules?

Relevant modules

Unification of similar modules

- Identify sets $\mathcal{M} = \{M_1, \ldots, M_k\}$ of modules that differ in only few axioms
- Replace $\mathcal{M}$ with $\bigcup \mathcal{M}$ and $\bigcap \mathcal{M}$

Outcome for Koala: no significant reduction in module numbers.
Unification of similar modules

- Identify sets $M = \{ M_1, \ldots, M_k \}$ of modules that differ in only few axioms
- Replace $M$ with $\bigcup M$ and $\bigcap M$
Unification of similar modules

- Identify sets $\mathcal{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_k\}$ of modules that differ in only few axioms.
- Replace $\mathcal{M}$ with $\bigcup \mathcal{M}$ and $\bigcap \mathcal{M}$.

Outcome for **Koala**: no significant reduction in module numbers.
Identify modules that are an agglomeration of other modules and contain no new information.

\( \mathcal{M} \) is fake if there is partition \( \mathcal{M} = \mathcal{M}_1 \uplus \cdots \uplus \mathcal{M}_k \) with pairwise disjoint sig(\( \mathcal{M}_i \)).

All other modules are genuine.
Genuine and fake modules

- Identify modules that are an agglomeration of other modules and contain no new information.

- \( M \) is \textit{fake} if there is partition \( M = M_1 \uplus \cdots \uplus M_k \) with pairwise disjoint \( \text{sig}(M_i) \).

\[
\begin{array}{c|c|c|c}
M_1 & M_2 & \cdots & M_k \\
\end{array}
\]

- All other modules are \textit{genuine}.

- Outcome for \textbf{Koala}:
  66% of the 3660 modules are genuine.
Weight analysis

... by scalesperson Chiara ;-)
Weight analysis

- Number of terms in the module \( m \)
Weight analysis

- Number of terms in the module \( m \)
- Minimal size of seed signatures \( s \)

\[ \Sigma_1 \]

\[ \text{sig}(\mathcal{M}) \]
Weight analysis

- Number of terms in the module: $m$
- Minimal size of seed signatures: $s$
- Number of different minimal seed signatures: $r$

$$
\text{Pulling Power } (M) = m \cdot s \cdot r
$$

$$
\text{Cohesion } (M) = \sum_{i=1}^{r} \Sigma_i
$$

$$
\text{Weight } (M) = w = r \cdot m \cdot s^2
$$
Weight analysis

- Number of terms in the module \( m \)
- Minimal size of seed signatures \( s \)
- Number of different minimal seed signatures \( r \)

\[ \sim \text{PullingPower}(\mathcal{M}) \]
What is my ontology about? How many modules? Relevant modules

**Weight analysis**

- Number of terms in the module $m$
- Minimal size of seed signatures $s$
- Number of different minimal seed signatures $r$

$\sim$ PullingPower($\mathcal{M}$) $\frac{m}{s}$

![Diagram](sig(\mathcal{M}) \rightarrow \Sigma_1, \Sigma_2, \ldots, \Sigma_r)$
Weight analysis

- Number of terms in the module \( m \)
- Minimal size of seed signatures \( s \)
- Number of different minimal seed signatures \( r \)

\[ \text{PullingPower}(\mathcal{M}) = \frac{m}{s} \]

\[ \text{Cohesion}(\mathcal{M}) \]
Weight analysis

- Number of terms in the module \( m \)
- Minimal size of seed signatures \( s \)
- Number of different minimal seed signatures \( r \)

\[
\begin{align*}
\text{PullingPower}(\mathcal{M}) & = \frac{m}{s} \\
\text{Cohesion}(\mathcal{M}) & = \frac{r}{s}
\end{align*}
\]
Weight analysis

- Number of terms in the module \( m \)
- Minimal size of seed signatures \( s \)
- Number of different minimal seed signatures \( r \)

\[
\text{Pulling Power}(\mathcal{M}) = \frac{m}{s}
\]

\[
\text{Cohesion}(\mathcal{M}) = \frac{r}{s}
\]

\[
\text{Weight}(\mathcal{M}) = w = \frac{r \cdot m}{s^2}
\]
Example: Koala

Animal ⊑ ≥ 1 hasHabitat
Animal ⊑ = 1 hasGender
DryEucalyptForest ⊑ Forest
Female ≡ ∃ hasGender. {female}
Forest ⊑ Habitat
GraduateStudent ⊑ Student
GraduateStudent ⊑ ∃ hasDegree. ({BA} ∪ {BS})
Koala ⊑ ∃ hasHabitat. DryEucalyptForest
Koala ⊑ Marsupials
Koala ⊑ ∃ isHardworking. {false}
KoalaWithPhD ≡ Koala ⊓ ∃ hasDegree. {PhD}
Male ≡ ∃ hasGender. {male}
Marsupials ⊑ Animal
Marsupials ⊑ ¬ Person
Parent ≡ Animal ⊓ ≥ 1 hasChildren
Parent ⊑ Animal
Person ⊑ Animal
Person ⊑ ¬ Marsupials
Quokka ⊑ ∃ isHardworking. {true}
Quokka ⊑ Marsupials
Rainforest ⊑ Forest
TasmanianDevil ⊑ Marsupials
University ⊑ Habitat
∃ hasChildren ⊑ Animal
∪ ⊑ ∀ hasChildren . Animal
∃ hasDegree ⊑ Person
MaleStudentWith3Daughters ≡ Student ⊓ ∀ hasChildren . Female ⊓ ∃ hasGender . {male} ⊓ = 1 hasChildren
Student ≡ Person ⊓ ∃ hasHabitat . University ⊓ ∃ isHardworking . {true}
Example: Koala

MaleStudentWith3Daughters, isHardWorking, University, Student, Parent, hasChildren

17 axioms

{Student, Parent}
{Student, hasChildren}
{MaleStudentWith3Daughters}
{hasChildren, University, isHardWorking}
{Parent, University, isHardWorking}
Example: Koala

- Koala, hasDegree, KoalaWithPhD
  - 7 axioms

- MaleStudentWith3Daughters, isHardWorking, University, Student, Parent, hasChildren
  - 17 axioms
Example: **Koala**

- Koala, hasDegree, KoalaWithPhD
  - 7 axioms

- MaleStudentWith3Daughters, isHardWorking, University, Student, Parent, hasChildren
  - 10 axioms

- Male, Female, hasHabitat, Animal, hasGender
  - 7 axioms
Example: Koala

After the first 12 heaviest modules . . .
Example: Koala

- Habitat
- Gender
- Forest
- Koala, hasDegree, KoalaWithPhD
- isHardWorking
- hasHabitat, hasGender, Animal

- Quokka
- Tasmanian Devil
- Degree

- Male Student With 3 Daughters
- hasChildren, Parent
- Student, University

- Male, Female, hasHabitat, hasGender

What is my ontology about? How many modules? Relevant modules
Work in progress

Is it necessary for the weight analysis to compute all modules?
Work in progress

Is it necessary for the weight analysis to compute all modules?

Remember: \[ \text{Weight}(\mathcal{M}) \quad w = \frac{r \cdot m}{s^2} \]

\[ \sim \text{search within modules with very small seed signatures} \]
What is my ontology about? How many modules? Relevant modules

Work in progress

Preliminary picture for **Periodic:**

```
memberOf / hasComponent / Atom
Ion / hasCharge
*Structure *Colour
MetalAtom / NonMetalAtom
Alkal*Ion / Cation
MoleOf* Molibdenum Tungstenum
NonMetal Metalloid
Datatype hasSolubilityIn Water
hasMolar Mass hasSolubilityIn Water Datatype
memberOf / hasComponent / Atom
*Structure *Colour
```
Outlook

- Find heaviest modules without computing all modules
- How many modules can ontologies have?
- Relation module number ↔ justificatory structure
What is my ontology about? How many modules? Relevant modules

Outlook

- Find heaviest modules without computing all modules
- How many modules can ontologies have?
- Relation module number ↔ justificatory structure

Thank you.