

Die modulare Struktur einer Ontologie: Atomare Dekomposition

Chiara Del Vescovo¹ Bijan Parsia¹
Uli Sattler¹ *Thomas Schneider*²

¹School of Computer Science, University of Manchester, GB

²Fachbereich Informatik, Universität Bremen

LogInf, 4. November 2011

... sind endliche Theorien in einer (Beschreibungs-)Logik, z. B.

$$\begin{aligned} \text{Part} &\equiv \exists \text{strict_part_of}.\text{Whole} \\ \text{strict_part_of} &\sqsubseteq \text{part_of} \\ \text{Hand} &\sqsubseteq =5\text{part_of}.\text{Finger} \end{aligned}$$

OWL (W3C-Standard, FOL-Fragment): expressive DL mit

- unären und binären Predikaten
- Konstruktoren für unäre Prädikate
 - Boolesch, \exists , \forall , Zählquantoren, Nominale
- Konstruktoren für binäre Prädikate
 - Inverses, Komposition
- Axiome: Inklusionen/Äquivalenzen von Prädikaten
- globalen Beschränkungen für Entscheidbarkeit

Module

Ein Modul $\mathcal{M}(\Sigma, \mathcal{O}) \subseteq \mathcal{O}$ kapselt Wissen
bezüglich einer Signatur Σ wenn $\mathcal{M} \equiv_{\Sigma}^{\mathcal{C}} \mathcal{O}$

d. h. für alle $C \sqsubseteq D$ mit $\text{sig}(C \sqsubseteq D) \subseteq \Sigma$:
 $\mathcal{O} \models C \sqsubseteq D$ gdw. $\mathcal{M}(\Sigma, \mathcal{O}) \models C \sqsubseteq D$

Bsp.: $\mathcal{O} = \text{Mereology.owl}$, $\Sigma = \{\text{Part, Whole, strict_part_of}\}$,
 $\mathcal{M}(\Sigma, \mathcal{O}) =$

	\mathcal{O}
	\mathcal{M}
Part $\equiv \exists \text{strict_part_of. Whole}$	strict_part_of $\equiv \text{direct_part_of}$
Part $\sqsubseteq \forall \text{strict_part_of. Whole}$	direct_part_of $\equiv \text{direct_part}^{-1}$
Whole $\equiv \exists \text{strict_part. Part}$	strict_part $\equiv \text{direct_part}$
Whole $\sqsubseteq \forall \text{strict_part. Part}$	strict_part_of $\equiv \text{strict_part}^{-1}$

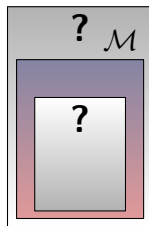
Modulare *Struktur*

Module sind prima:

- wenn man die nötige (Start-)Signatur kennt ...
- für „Ein-Modul“-Aufgaben wie Wiederverwendung

Extraktion *eines* Moduls hilft *nicht*, wenn man

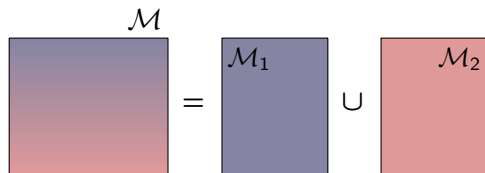
- die *richtige* Startsignatur *nicht* kennt
- *andere* Module verstehen möchte
- die *Abhängigkeitsstruktur der Axiome* verstehen möchte



Zum Analysieren der modularen Struktur der Ontologie:

- **signifikante** Module
- **signifikante** Relationen zwischen Modulen
- ... offenbart logische Abhängigkeiten zwischen Axiomen

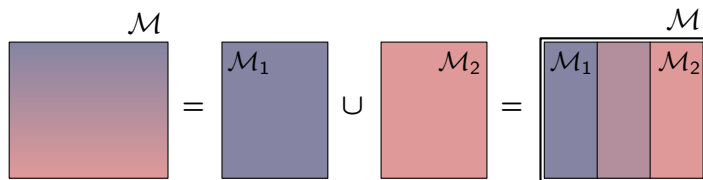
Sind alle Module signifikant?



Um \mathcal{M} zu verstehen, muss man verstehen:

- die Abhängigkeitsstruktur von \mathcal{M}_1
- die Abhängigkeitsstruktur von \mathcal{M}_2
- nichts weiter: \mathcal{M}_1 und \mathcal{M}_2 haben keine weiteren Abhängigkeiten

Sind alle Module signifikant?



Um \mathcal{M} zu verstehen, muss man verstehen:

- die Abhängigkeitsstruktur von \mathcal{M}_1
- die Abhängigkeitsstruktur von \mathcal{M}_2
- nichts weiter: \mathcal{M}_1 und \mathcal{M}_2 haben keine weiteren Abhängigkeiten

↪ \mathcal{M} ist nicht signifikant: es ist ein **unechtes** Modul

- \mathcal{M}_1 und \mathcal{M}_2 können signifikant sein.
- Es ist wichtig zu wissen, dass \mathcal{M} „nur“ eine Vereinigung ist.

Man betrachte ein Modul \mathcal{M} , das **nicht unecht** ist.

Um \mathcal{M} zu verstehen, muss man \mathcal{M} **im Ganzen** verstehen.

- alle Axiome in \mathcal{M} interagieren logisch
- möglicherweise auf verschiedene Weisen

„Nicht unecht“ impliziert signifikant: **echtes** Modul



Verhältnis echt zu unecht

Gegeben eine Menge echter Module,

- Vereinigungen führen zu unechten Modulen

↪ exponentiell großer Raum für unechte Module

- Aber: nicht *jede* Vereinigung echter Module ist ein Modul

Anzahl *aller* Module wächst offenbar exponentiell in $|\mathcal{O}|$

[D.,P.,S.,S., KR 2010 & WoMO 2010]

Frage 1

Kommt hohe Modulanzahl primär durch triviale Kombinationen zustande?

D. h., sind die meisten Module **unecht**?

Theorem 1

Jedes echte Modul ist das kleinste Modul, das α enthält, für ein Axiom $\alpha \in \mathcal{O}$.

↪ Familie der echten Module ist linear in $|\mathcal{O}|$.

Die meisten Module sind unecht!

Beweis nutzt Eigenschaften von Modulen aus:

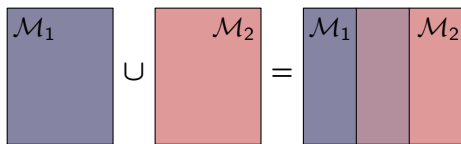
- Eindeutigkeit, Monotonie, Abgeschlossenheit, ...
- werden von allen lokalitätsbasierten Modulen erfüllt

Relationen zwischen Modulen

Echte Module können sich überlappen.

Das zeigt signifikante logische Abhängigkeiten zwischen Axiomen auf:

Die Axiome in $\mathcal{M}_1 \setminus \mathcal{M}_2$ hängen von den Axiomen in $\mathcal{M}_1 \cap \mathcal{M}_2$ ab.

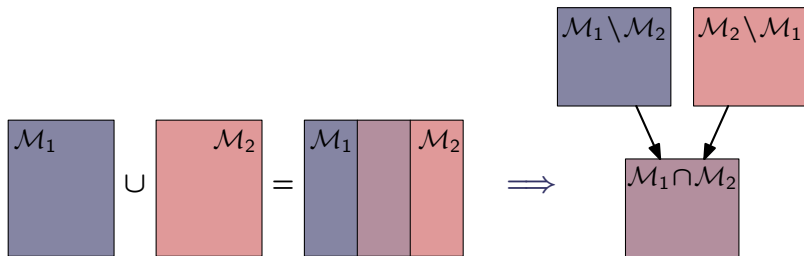


Relationen zwischen Modulen

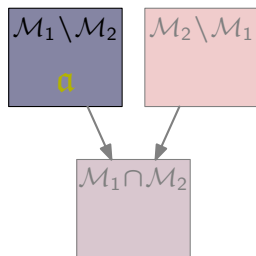
Echte Module können sich überlappen.

Das zeigt signifikante logische Abhängigkeiten zwischen Axiomen auf:

Die Axiome in $\mathcal{M}_1 \setminus \mathcal{M}_2$ hängen von den Axiomen in $\mathcal{M}_1 \cap \mathcal{M}_2$ ab.



Atom = maximale Menge $\alpha \subseteq \mathcal{O}$, so dass für jedes Modul \mathcal{M} :
entweder $\alpha \subseteq \mathcal{M}$ oder $\alpha \cap \mathcal{M} = \emptyset$.



- Das kleinste Modul für ein Axiom α enthält das gesamte Atom mit α .
 - Axiome in einem Atom sind logisch voneinander abhängig.
 - Atome sind paarweise disjunkt.
- ↪ Die Familie aller Atome partitioniert die Ontologie.
- Jedes Modul ist eine disjunkte Vereinigung von Atomen.

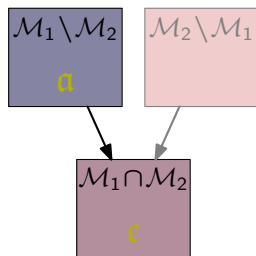
Proposition

Es gibt 1-1-Korrespondenz zwischen echten Modulen und Atomen.

Atomare Dekomposition

Abhängigkeit zwischen Atomen:

- $\alpha \succeq \mathfrak{c}$ gdw. für alle \mathcal{M} :
wenn $\alpha \subseteq \mathcal{M}$, dann $\mathfrak{c} \subseteq \mathcal{M}$
- Axiome in α hängen logisch von Axiomen in \mathfrak{c} ab



Theorem 2

Die Relation \succeq ist reflexiv, antisymmetrisch und transitiv.

\rightsquigarrow Hasse-Diagramm stellt 2 logische Abhängigkeiten dar:
zwischen Axiomen in Atomen, zwischen Atomen

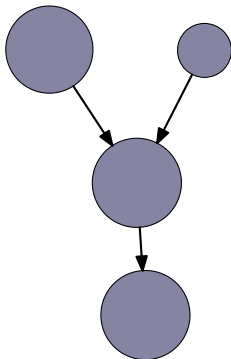
42 Axiome
1952 Module

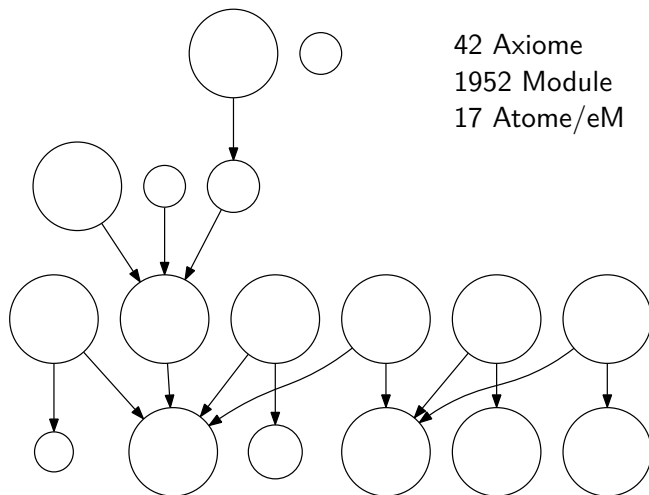


42 Axiome
1952 Module



42 Axiome
1952 Module





Frage 2

Kann man

- alle echten Module berechnen,
- alle Atome berechnen,
- die Abhängigkeiten zwischen Atomen berechnen,

ohne alle Module zu berechnen?!

Zur Erinnerung:

Theorem 1

Jedes echte Modul ist das kleinste Modul, das α enthält,
für ein Axiom $\alpha \in \mathcal{O}$.

- Extrahiere $\mathcal{M}(\text{sig}(\alpha), \mathcal{O})$
(\leq linear viele Modulextraktionen)
- Atomare Dekomposition wird bereits durch Vergleich nur der
echten Module induziert
(quadratische Prozedur)

Haben atomare Dekomposition auf 181 OWL-Ontologien aus dem NCBO BioPortal angewandt.

Zerlegbarkeit:

Mittlere Anz. Axiome pro Atom	1.73
" max. Anz. Axiome pro Atom	86
" Anz. Axiome pro echtes Modul	66
" max. Anz. Axiome pro echtes Modul	143

- Atomare Dekomposition (AD) ist lineare Repräsentation der potentiell exponentiellen Menge aller Module.
- AD kann mittels einer linearen Anzahl von Modulextraktionen berechnet werden.
- AD zeigt 2 Arten logischer Abhängigkeiten zwischen Axiomen auf.

- Abhängigkeiten zwischen Atomen und *Mengen* von Atomen
- Beschriftungen für Atome –
verschiedene Beschriftungen für verschiedene Aufgaben
- Anwendungen
 - „Themen“ für das Verstehen von Ontologien
 - Schnelle Modulextraktion
 - Bestimmung der Modulanzahl
 - ...

- Abhängigkeiten zwischen Atomen und *Mengen* von Atomen
- Beschriftungen für Atome –
verschiedene Beschriftungen für verschiedene Aufgaben
- Anwendungen
 - „Themen“ für das Verstehen von Ontologien
 - Schnelle Modulextraktion
 - Bestimmung der Modulanzahl
 - ...

Vielen Dank.

Ontologien mit problematischer Zerlegbarkeit

Ontology \mathcal{O} (ID in BioPortal)	$\#\mathcal{O}$	$\#\max$ Atom	$\#\text{Eq.}$ axs.	$\#\text{Disj.}$ 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