Ontology Partitioning Using $\mathcal{E}$-Connections Revisited

Sascha Jongebloed and *Thomas Schneider*

University of Bremen

DL 2018 Tempe, AZ, USA Oct 28

Introduction

Modularity

Large ontologies with 100,000s of axioms

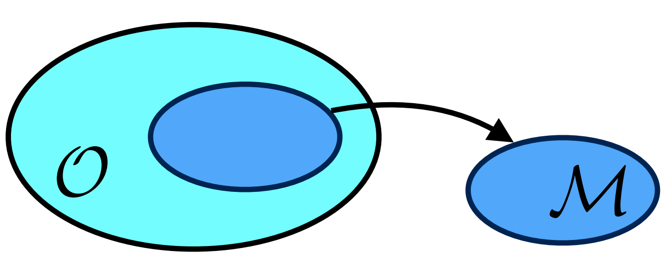
e.g. SNOMED CT The global language of healthcare   NATIONAL CANCER INSTITUTE

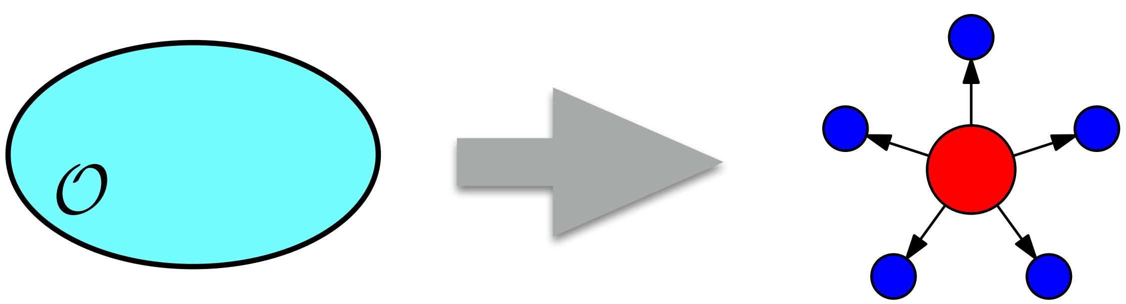
Challenges

- Loading, navigation
- Understanding the logical structure (comprehension)
- Efficient automated reasoning
- Efficient re-use
- Versioning   and more …

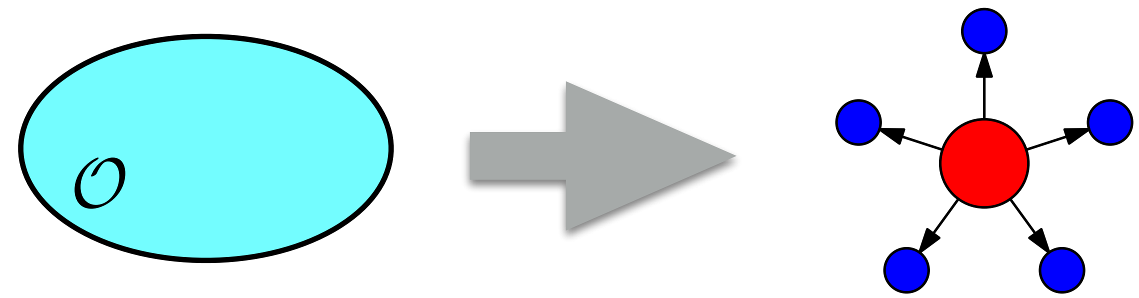
Modularity helps:

Module extraction   and   **Decomposition**





Decomposition



Existing approaches

- Signature splitting [Parikh '99]
- Signature Δ-decomposition [Konev et al. '10]

Partitions based on $\mathcal{E}$-connections [Cuenca Grau et al. '06]

- Atomic decomposition [Del Vescovo et al. '11]
- Structure-based partitioning
  [Stuckenschmidt & Klein '04, Amato et al. '15]

$\mathcal{E}$-Partitions in a Nutshell

Aim: Automatic and efficient partitioning of an ontology; parts are connected via “semantic links” in the style of $\mathcal{E}$-connections

$\mathcal{E}$-connections . . . [Kutz et al. 2004]

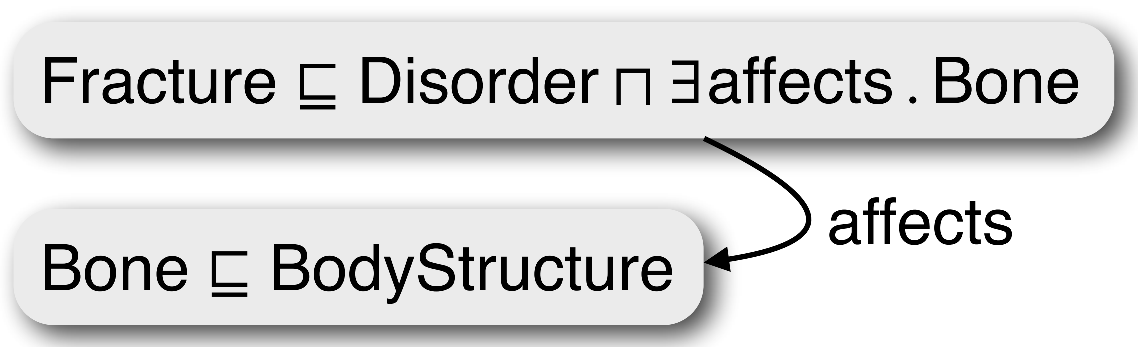
- combine (heterogeneous) logical theories via *link relations*
- semantics via partitioned interpretations

An $\mathcal{E}$-partition of an ontology $\mathcal{O}$ . . . [Cuenca Grau et al. 2006]

- is the unique maximal $\mathcal{E}$-connection equivalent to $\mathcal{O}$ (with link relations from $\mathcal{O}$’s role names)
- can be computed in polytime for $\mathcal{O}$ in DLs up to $\mathcal{SHOIQ}(\mathcal{D})$
- its components are logically encapsulating

e.g.:

$\text{Fracture} \sqsubseteq \text{Disorder} \sqcap \exists \text{affects} . \text{Bone}$
$\text{Bone} \sqsubseteq \text{BodyStructure}$



Our Aims

Where we started

Understand algorithm?

Fix bugs in original implementation?

Where we got

- found a simpler algorithm that runs in linear time
- simplified notation and proofs
- extended the approach to (almost) OWL
- identified potential for extension beyond OWL and limits

Work in progress!

$\mathcal{E}$-Connections and $\mathcal{E}$-Partitions for OWL

Indexing the Vocabulary

Let S be an arbitrary index set.

Index function ι

- (concept names) $A \mapsto$ index $\iota(A) \in S$
  (role names) $r \mapsto$ pair of indices $\iota(r)$
- is extended to complex concepts:

  e.g., $\iota(\exists r.D) = i$ if $\iota(r) = (i, j)$ and $\iota(D) = j$

  (and many more cases)

  $\exists r.D$ is ι-wellformed
- is extended to axioms:

  e.g., $\iota(C \sqsubseteq D) = i$ if $\iota(C) = \iota(D) = i$
- and thus determines a partitioning of ontologies

Two views on an ontology:

- as a monolithic ontology
- as a ι-ontology – an $\mathcal{E}$-connection!

Semantics of $\mathcal{E}$-Connections

ι-interpretations

- Domain $\Delta^{\mathcal{I}}$ is partitioned into $(\Delta_i^{\mathcal{I}})_{i \in S}$
- Concept names A with $\iota(A) = i$ are interpreted within $\Delta_i^{\mathcal{I}}$, analogously for role names
- Extension to complex concepts as usual
  except negation: $(\neg C)^{\mathcal{I},\iota} = \Delta^{\mathcal{I}}_{\underline{\boldsymbol{\iota(C)}}} \setminus C^{\mathcal{I},\iota}$

Two views on semantics:

- Standard semantics, denoted $\mathcal{I} \models \mathcal{O}$
- Semantics w.r.t. indexing ι, denoted $\mathcal{I} \models^{\iota} \mathcal{O}$

Compatibility and Equivalence

Let $\mathcal{O}$ be an ontology and $\mathbb{O} = (\mathcal{O}_i)_{i \in S}$ a ι-ontology.

Important relationships between $\mathcal{O}$ and $\mathbb{O}$:

- $\mathcal{O}$ and $\mathbb{O}$ are **compatible**, written $\mathcal{O} \sim \mathbb{O}$, if $\mathcal{O} = \biguplus_{i \in S} \mathcal{O}_i$.
- $\mathcal{O}$ and $\mathbb{O}$ are **equivalent**, written $\mathcal{O} \equiv \mathbb{O}$, if for all ι-interpret. $\mathcal{I}$:

$$\mathcal{I} \models \mathcal{O} \quad \text{iff} \quad \mathcal{I} \models^{\iota} \mathbb{O}$$

Apparently, compatibility and equivalence do **not** imply each other!

Domain-Independence

Well-known notion from database theory relates compatibility & equivalence:

$\mathcal{O}$ is **domain-independent (DI)**
if for all interpretations $\mathcal{I}, \mathcal{J}$ with $X^{\mathcal{I}} = X^{\mathcal{J}}$ for all terms X:

$$\mathcal{I} \models \mathcal{O} \quad \text{iff} \quad \mathcal{J} \models \mathcal{O}$$

Nice characterization of all DI concepts [Cuenca Grau et al. 2006]
allows to check DI in linear time; additionally gives:

- If C is not DI and $\mathcal{I}, \mathcal{J}$ are as above with $\Delta^{\mathcal{J}} = \Delta^{\mathcal{I}} \uplus S$, then $C^{\mathcal{J}} = C^{\mathcal{I}} \cup S$.

Holds for all of OWL except the universal role.

Domain-Independence

Previous characterization is crucial in the proof of the following:

Theorem.

1. If $\mathcal{O}$ is DI and $\mathcal{O} \sim \mathbb{O}$, then $\mathcal{O} \approx \mathbb{O}$.
2. If additionally $\mathcal{O}$ is consistent, then so is $\mathbb{O}$.

Consequence

For DI ontologies,
it suffices to compute the minimal compatible E-connection.

- From now on, we assume that the input ontology $\mathcal{O}$ is DI.

The New Partitioning Algorithm and First Tests

A Simple Algorithm

Idea:

For input ontology $\mathcal{O}$,

find index set S of maximal cardinality and index function ι such that all concepts and axioms in $\mathcal{O}$ are ι-wellformed

The Algorithm:

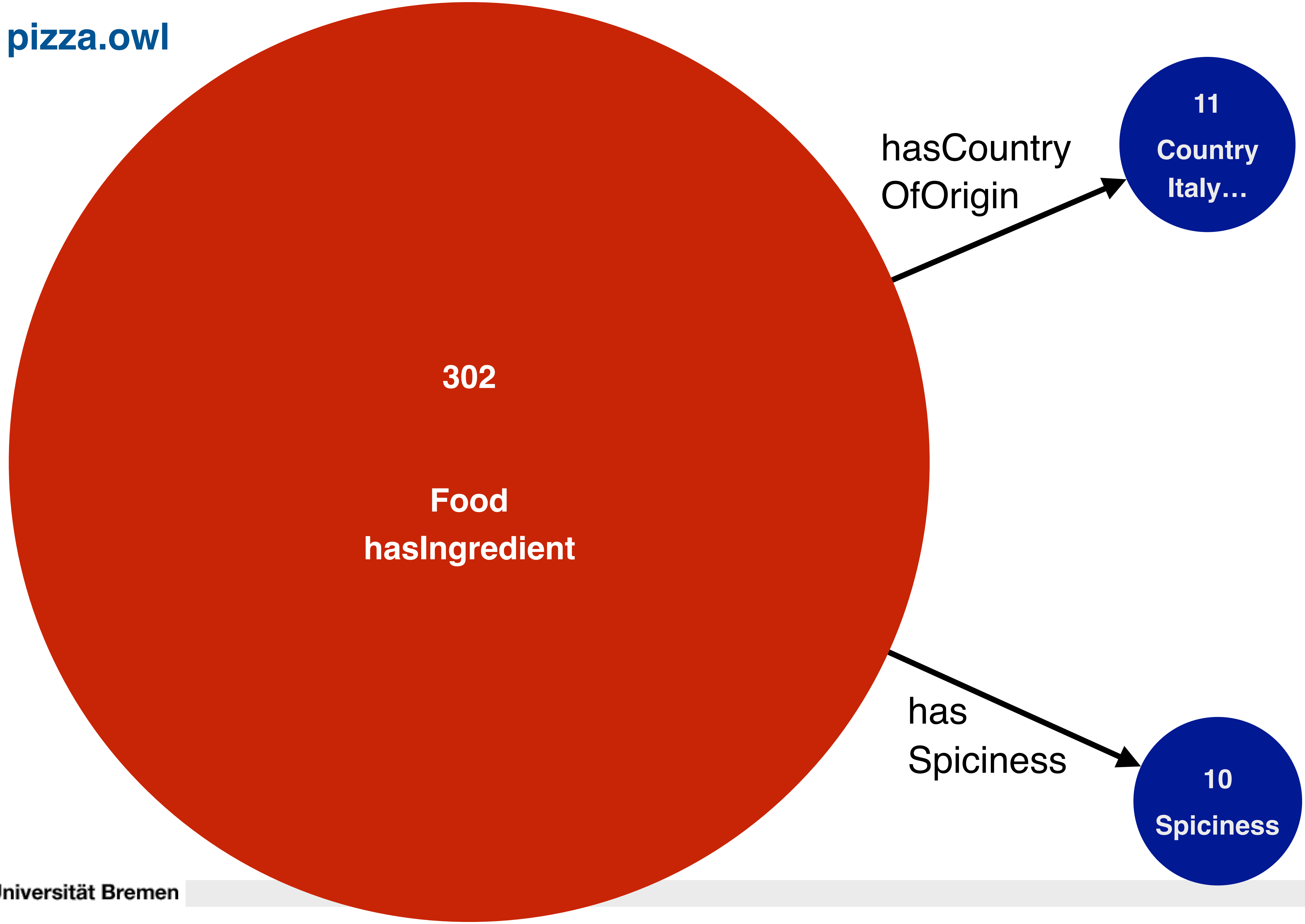
1. Collect wellformedness constraints in an undirected graph G
   - nodes: one per (complex) concept, 2 per role name
   - edges = constraints
2. G's connected components induce S, ι, $\mathbb{O}$

Both steps easy to implement in **linear time.**

Correctness and maximality are straightforward to show:
algo mimics wellformedness definition!

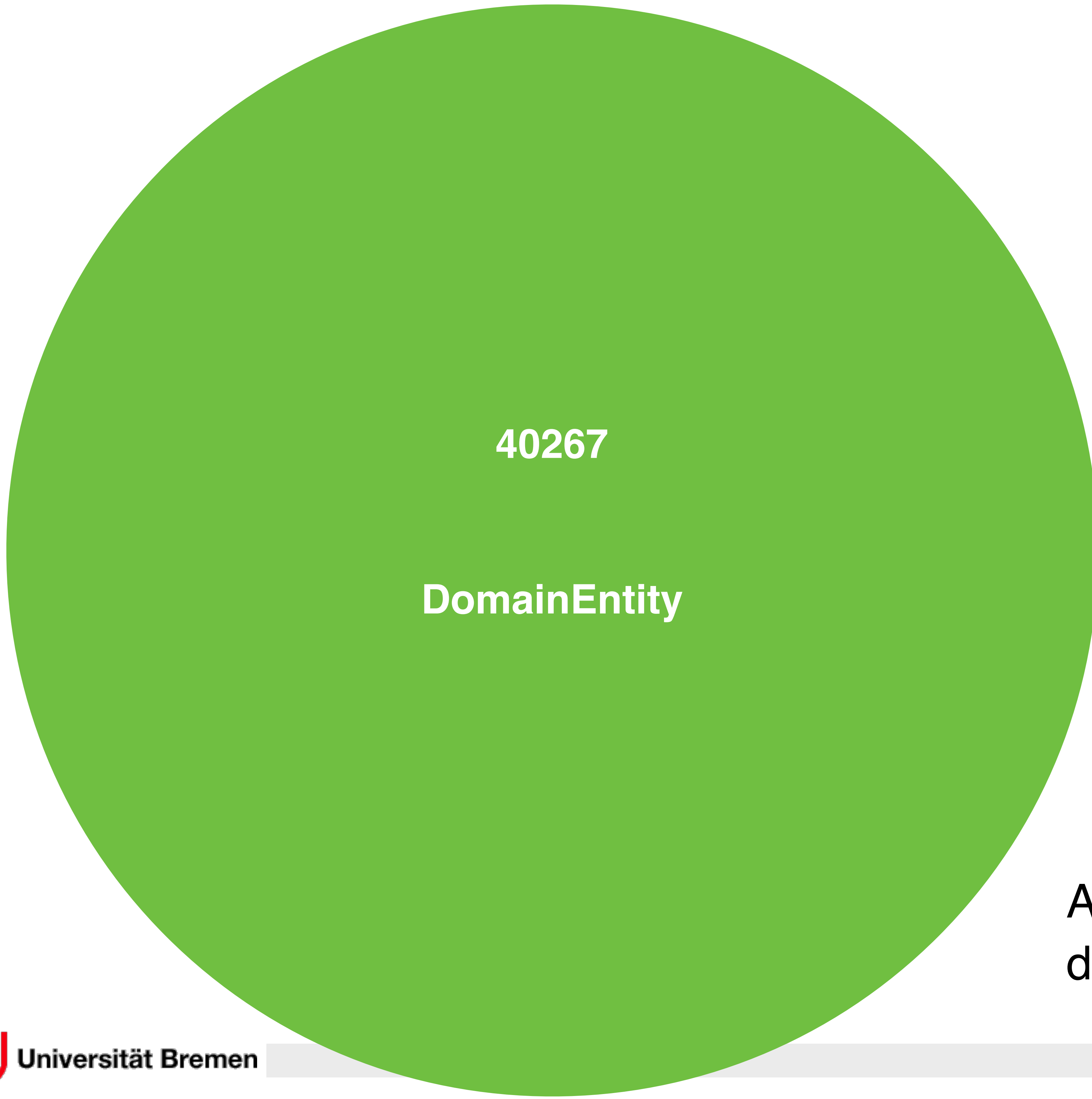
Example Decomposition: Pizza Ont.

pizza.owl



Example Decomposition: PTO

Periodic Table Ontology by Robert Stevens



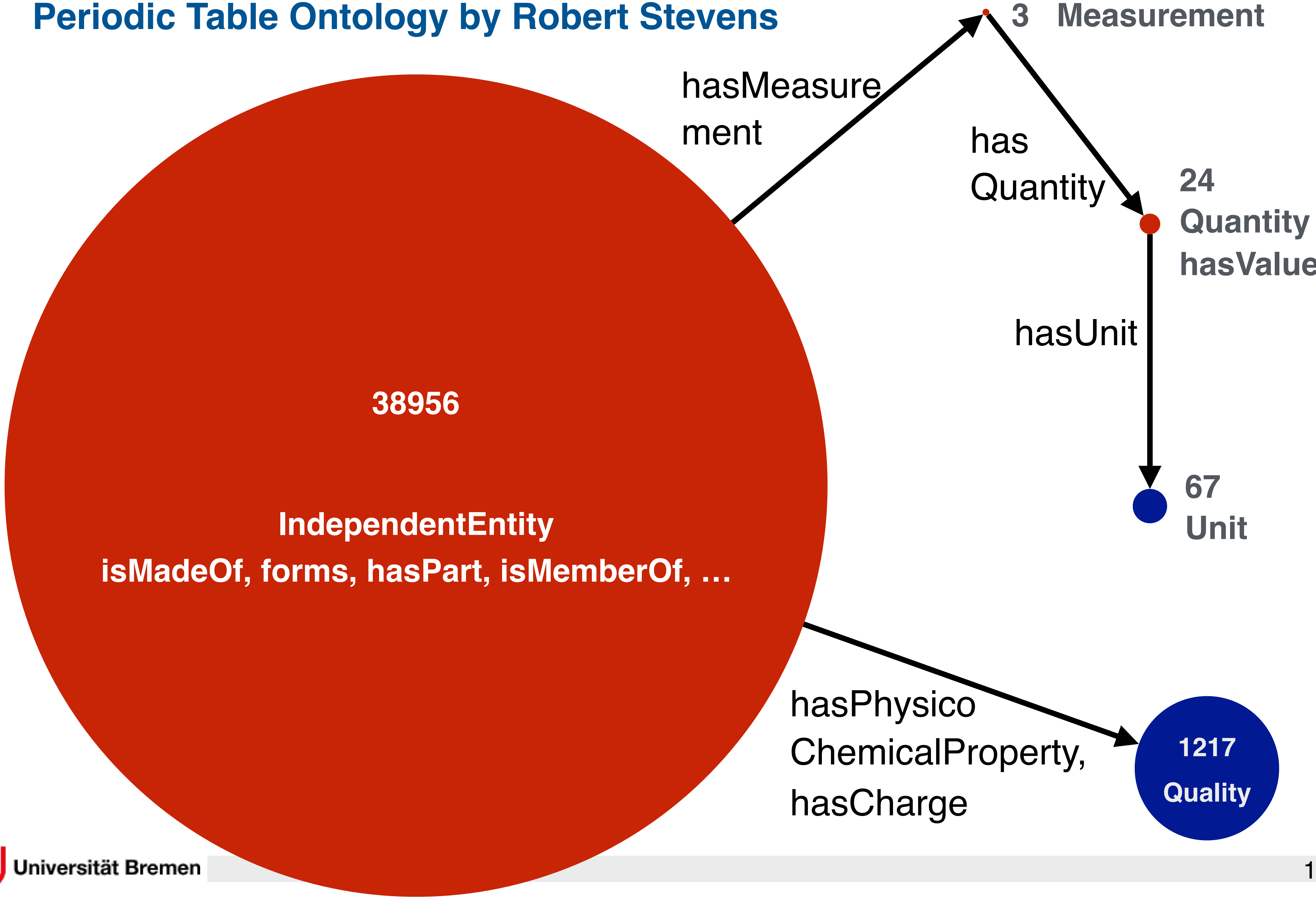
Top-level concepts forbid decomposition

Heuristics:
delete top *n* levels

Alternatively,
delete **upper-level concepts**

Decomposition: PTO with 3 levels removed

Periodic Table Ontology by Robert Stevens



Outlook

Outlook

Coming soon:

- Systematic evaluation
- Heuristics for ontologies that don’t decompose well
- Extensions: TGDs, UNFO?

The End

Questions?

¿Preguntas?

Fragen?

Vragen?

Pytania?

Thank you.

Kysymyksiä?

Vrae?

Ερωτήσεις;

Întrebări?

Вопросы?

Questões?