Combining Concept Annotation and Pattern Structures for Guiding Ontology Mapping

Pierre Monnin[1], Amedeo Napoli[1] and Adrien Coulet[1,2]

July 13, 2018

[1] LORIA (CNRS, Inria Nancy-Grand Est, Université de Lorraine), Nancy, France
[2] BMIR (Stanford University), Stanford, USA

Context & Motivation

Context & Motivation (1/6)

Ontology

- A formal representation of a particular domain
- Composed of a *terminological component* – TBox
  $\rightarrow$ Definition of classes and predicates between classes
- Composed of an *assertion component* – ABox
  $\rightarrow$ Individuals instantiating classes and predicates

Example

The drug *codein* is an individual instantiating the class *Analgesics*

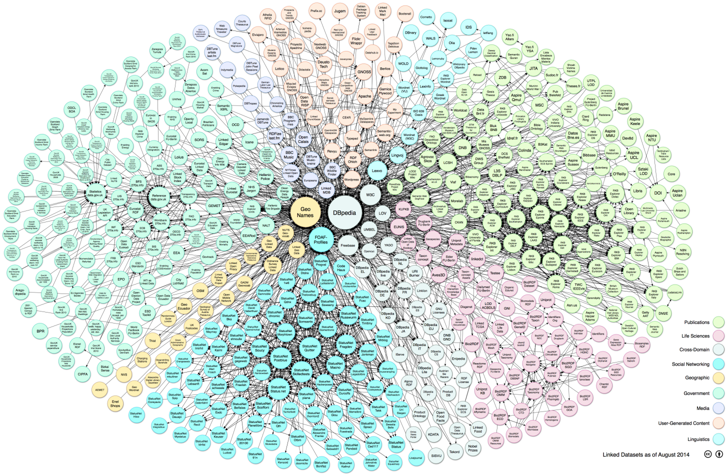


Figure 1: Linked Open Data cloud diagram 2014-08-30, by Andrejs Abele, Paul Buitelaar, Richard Cyganiak, Anja Jentzsch and John P. McCrae. http://lod-cloud.net/

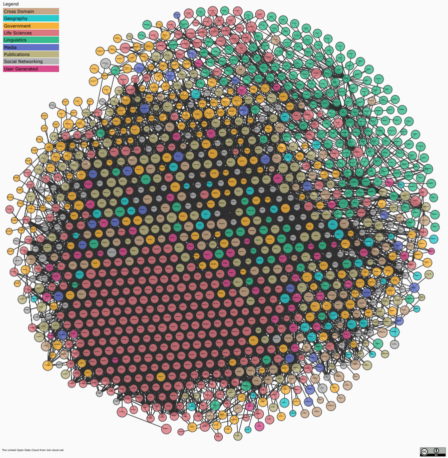


Figure 2: Linked Open Data cloud diagram 2018-06-28, by Andrejs Abele, Paul Buitelaar, Richard Cyganiak, Anja Jentzsch and John P. McCrae. http://lod-cloud.net/

Context & Motivation (4/6)

Use Case 1: There is a need for **structure between ontologies**

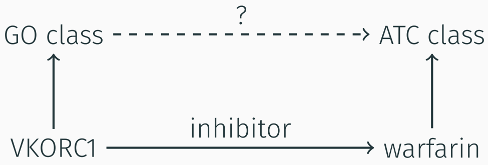
- Individuals can instantiate classes from several TBoxes
  E.g., in the medical domain, MeSH, ICD-9-CM, ICD-10-CM...
- Corresponding classes from several ontologies may be mapped with *equivalence relationships*
- This *alignment* may be a manual review by an expert or a semi-automatic process

Example
The class `Cardiomyopathies` (D009202) from MeSH is equivalent to `Cardiomyopathy` (425) in ICD-9-CM

Use Case 2: There is a need for **structure within an ontology**

- In the ABox: predicate assertions, *i.e.*, relations between individuals
  E.g., in DrugBank, *VKORC1* is an inhibitor of *warfarin*
- Individuals instantiate classes of ontologies
  E.g., GO, ATC
- From predicate assertions, frequently associated classes as domain and range could be discovered
  → Could indicate common behavior at the class level

Example



Context & Motivation (6/6)

Formal Concept Analysis (FCA) [Ganter and Wille, 1999]

- A well-fonded mathematical framework
- Already applied for knowledge engineering purposes

Pattern Structures [Ganter and Kuznetsov, 2001]

- An extension of FCA allowing to have complex descriptions of objects
  $\rightarrow$ Ontology classes

Concept Annotation [Monnin et al., 2017]

- An extension of FCA introduced in our previous work
- Adding a third dimension to a concept lattice w/o changing its structure
- Already experimented for ontology re-engineering

$\rightarrow$ Combine Concept Annotation and Pattern Structures for the two considered use-cases

Basics about Ontologies

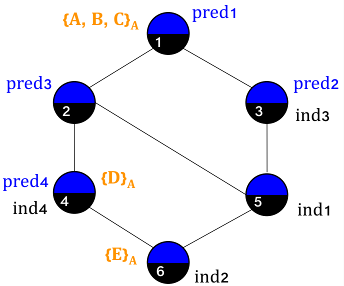
Basics about Ontologies

Considering an ontology $\mathcal{O}$

- We denote $\mathcal{C}(\mathcal{O})$ the set of classes of $\mathcal{O}$
- Subsumption relation $C \sqsubseteq D$
  $\rightarrow$ Partial order between classes
- $\mathrm{lcs}(C_1, C_2)$ is the *least common subsumer* of C_1 and C_2
- Considering a set of classes $\mathcal{C}_n = \{C_1, C_2, \ldots, C_n\}$
  $\rightarrow \min \mathcal{C}_n = \{C \in \mathcal{C}_n \mid \nexists\, D \in \mathcal{C}_n,\ D \sqsubseteq C\}$.

Concept Annotation

Concept Annotation: previous work



Formal context (G, M, I)

- G: set of **individuals**
- M: set of **predicates**
- $(g, m) \in I$ iff g is involved in a relationship whose predicate is m

Annotation

- Derivation operator $(\cdot)^\diamond : 2^G \to 2^{\mathcal{C}(\mathcal{O})}$
- Given a formal concept (A, B)
  $A^\diamond = \bigcap_{g \in A} \{C \mid \mathcal{O} \models C(g)\}$

$\rightarrow$ Annotated concept $(A, B, A^\diamond)$

Combining Concept Annotation with Pattern Structures

Objective: only keep as annotation the most specific classes instantiated

Mapping function δ

- $\delta : G \rightarrow 2^{\mathcal{C}(\mathcal{O})}$
- $\delta(g) = \min \{C \mid \mathcal{O} \models C(g)\}$

Similarity operator $\sqcap$

- Given $g_1, g_2 \in G$
- $\delta(g_1) \sqcap \delta(g_2) = \min \{\mathrm{lcs}(C_1, C_2) \mid \forall\, (C_1, C_2) \in \delta(g_1) \times \delta(g_2)\}$

Derivation operator $(\cdot)^\circ$

- $(\cdot)^\circ : 2^G \rightarrow 2^{\mathcal{C}(\mathcal{O})}$
- Given a formal concept (A, B), $A^\circ = \prod_{g \in A} \delta(g)$

Use Case 1: Suggesting Mappings between Classes of Ontologies

Suggesting Mappings between Classes of Ontologies (1/4)

Running Example

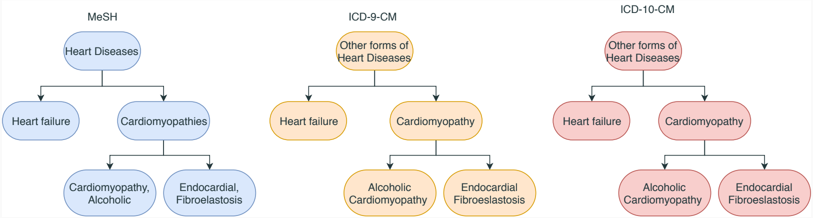
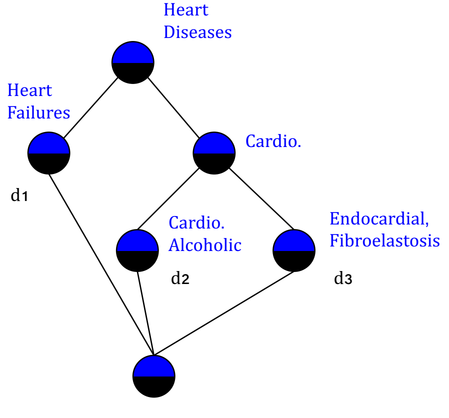


Figure 3: TBoxes

	$\mathcal{O}_{ref}$ (MeSH)	$\mathcal{O}_1$ (ICD-9-CM)	$\mathcal{O}_2$ (ICD-10-CM)
d1	Heart failure	Heart failure	Heart failure
d2	C., Alcoholic	Alcoholic C.	Alcoholic C.
d3	E., Fibroelastosis	E. Fibroelastosis	E. Fibroelastosis

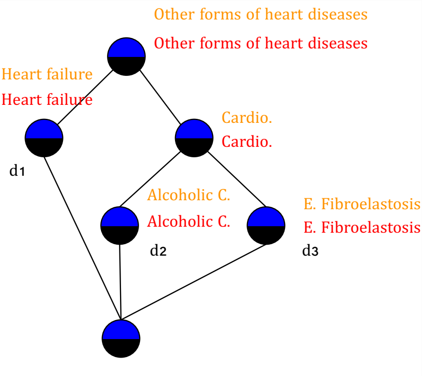
Table 1: Classes instantiated by the individuals (d1, d2, d3) of the ABox



Classifying individuals w.r.t. $\mathcal{O}_{ref}$

- Pattern Structure $(G, (2^{\mathcal{C}(\mathcal{O}_{ref})}, \sqcap_{ref}), \delta_{ref})$
- Pattern Concepts (A, D) with $A \subseteq G$ and $D \in 2^{\mathcal{C}(\mathcal{O}_{ref})}$

Suggesting Mappings between Classes of Ontologies (3/4)



Annotating with $\mathcal{O}_1$ and $\mathcal{O}_2$

- Two annotations for a formal concept (A, B)
  - w.r.t. $\mathcal{O}_1$

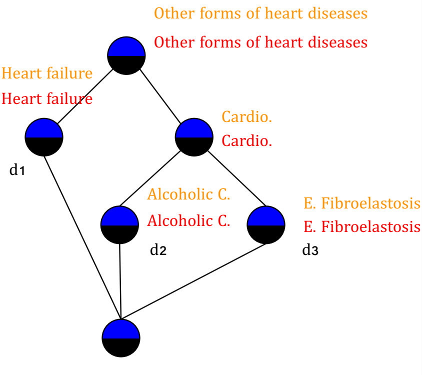
$$A^{\circ_1} = \prod_{g \in A} {}_1\, \delta_1(g)$$

  - w.r.t. $\mathcal{O}_2$

$$A^{\circ_2} = \prod_{g \in A} {}_2\, \delta_2(g)$$

$\rightarrow$ Annotated concept $(A, B, A^{\circ_1}, A^{\circ_2})$

Suggesting Mappings between Classes of Ontologies (4/4)



Reading mappings from the annotated lattice

Considering an annotated concept $(A, B, A^{\circ_1}, A^{\circ_2})$

- $A^{\circ_1} \subseteq \mathcal{C}(\mathcal{O}_1)$
- $A^{\circ_2} \subseteq \mathcal{C}(\mathcal{O}_2)$

Considering each pair $(C_1, C_2) \in A^{\circ_1} \times A^{\circ_2}$

- They are instantiated by all individuals in A
- Suggestion of $C_1 \Leftrightarrow C_2$

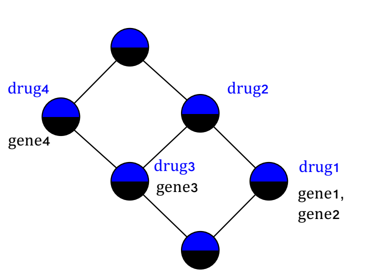
Use Case 2: Discovering Associated Classes as Domain and Range of a Predicate

Running example

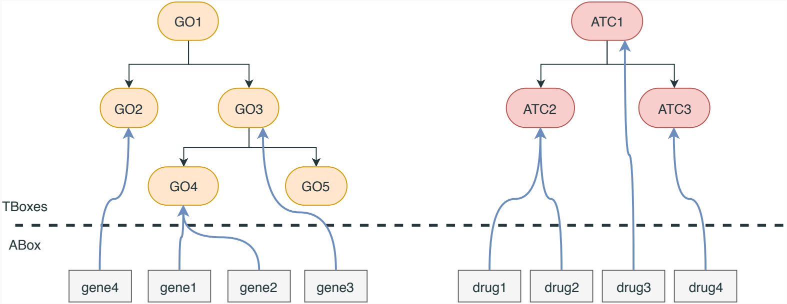
- A set G of individuals instantiating classes of an ontology $\mathcal{O}_1$
  $\rightarrow$ genes instantiating classes of GO
- A set M of individuals instantiating classes of an ontology $\mathcal{O}_2$
  $\rightarrow$ drugs instantiating classes of ATC
- The incidence relation $I \subseteq G \times M$ indicates that an individual from G is in a relationship with an individual from M
  $\rightarrow$ the gene is an inhibitor of the drug

Discovering Associated Classes as Domain and Range (2/5)

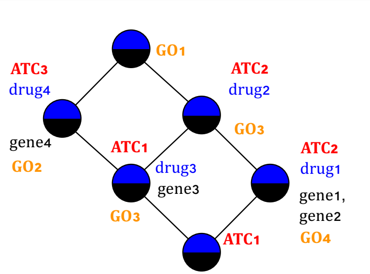
	$drug_1$	$drug_2$	$drug_3$	$drug_4$
$gene_1$	×	×		
$gene_2$	×	×		
$gene_3$		×	×	×
$gene_4$				×



Discovering Associated Classes as Domain and Range (3/5)



Discovering Associated Classes as Domain and Range (4/5)



Annotating with $\mathcal{O}_1$ and $\mathcal{O}_2$

- Two annotations for a formal concept (A, B)
  - w.r.t. $\mathcal{O}_1$

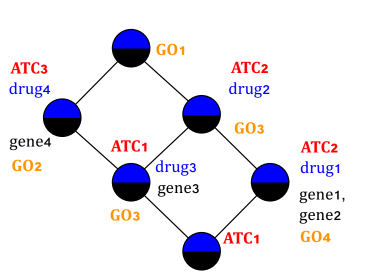
$$A^{\circ_1} = \textstyle\prod_{g \in A}^{1} \delta_1(g)$$

  - w.r.t. $\mathcal{O}_2$

$$B^{\circ_2} = \textstyle\prod_{m \in B}^{2} \delta_2(m)$$

$\rightarrow$ Annotated concept $(A, B, A^{\circ_1}, B^{\circ_2})$

Discovering Associated Classes as Domain and Range (5/5)



Reading Domain and Range from the annotated lattice

- Considering an annotated concept $(A, B, A^{\circ_1}, B^{\circ_2})$
- Every individual in A is in relationship w/ every individual in B
- $A^{\circ_1} \subseteq \mathcal{C}(\mathcal{O}_1)$ is the set of the most specific classes instantiated by all individuals in A
- $B^{\circ_2} \subseteq \mathcal{C}(\mathcal{O}_2)$ is the set of the most specific classes instantiated by all individuals in B

$\rightarrow$ Classes from A°_1} as domain are associated with classes from B°_2} as range

Conclusion & Perspectives

Conclusion & Perspectives

- Pattern structures enable more complex annotations
- The resulting lattice can be seen as a *pivot* structure
- The approach needs to be validated on real data sets
- Mappings suggestion
    - Closed World Assumption *vs* Open World Assumption
    - Choice of background knowledge $\mathcal{O}_{ref}$?
    - Potential other applications: concept drift...
- Discovering associated classes as domain and range
    - Interactive exploration of the annotated lattice
    - Add metrics (*e.g.*, *support*) for selecting interesting associations?

Thank you for your attention

Questions?

Ganter, B. and Kuznetsov, S. O. (2001).
Pattern Structures and Their Projections.
In *Conceptual Structures: Broadening the Base, 9th International Conference on Conceptual Structures, ICCS 2001, Stanford, CA, USA, July 30-August 3, 2001, Proceedings*, pages 129–142.

Ganter, B. and Wille, R. (1999).
Formal concept analysis - mathematical foundations.
Springer.

Monnin, P., Lezoche, M., Napoli, A., and Coulet, A. (2017).
Using Formal Concept Analysis for Checking the Structure of an Ontology in LOD: The Example of DBpedia.
In *Foundations of Intelligent Systems - 23rd International Symposium, ISMIS 2017, Warsaw, Poland, June 26-29, 2017, Proceedings*, pages 674–683.