Construction and exploitation of an historical knowledge graph to deal with the evolution of ontologies

Silvio Domingos Cardoso, Chantal Reynaud-Delaître, Marcos Da Silveira, **Cédric Pruski**

Journées RoCED, 6th-7th of July 2021

université PARIS-SACLAY

LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY | LIST

PROBLEMATIC

Ontology evolution management

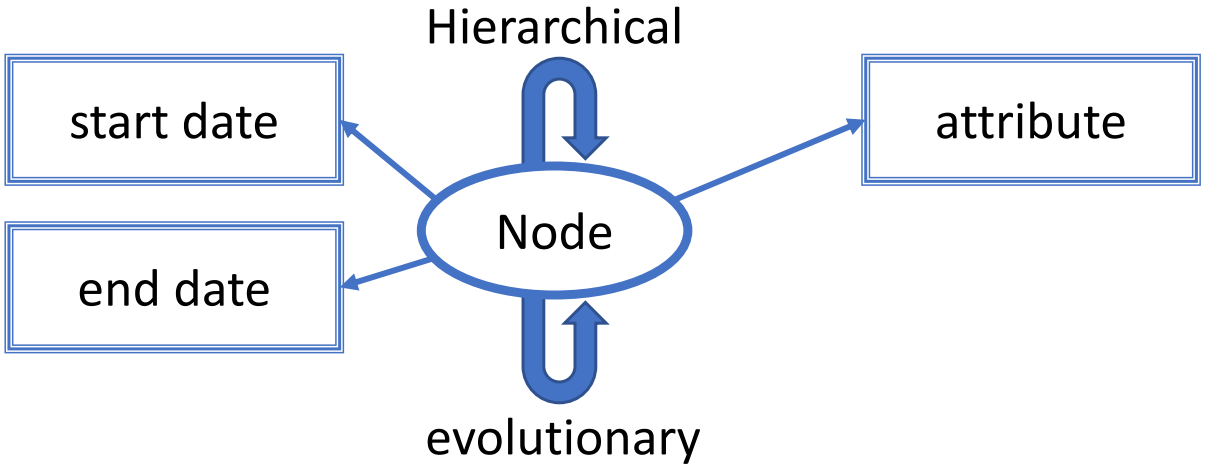
LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY | LIST

- More and more smart applications rely on ontologies for their modelling and reasoning capabilities
  - Ex: BBC, PubMed, Elsevier, Google, Amazon, Dynaccurate …
- But ontologies evolve over time mainly for knowledge evolution purposes
  → New ontology versions are regularly published
- Tools exist to access these various ontology versions but evolution links (between entities of successive versions) are never specified

We propose a Knowledge Graph that contains all versions of ontological concepts with an explicit specification of the evolution link between these concepts

HISTORICAL KNOWLEDGE GRAPH

The metamodel



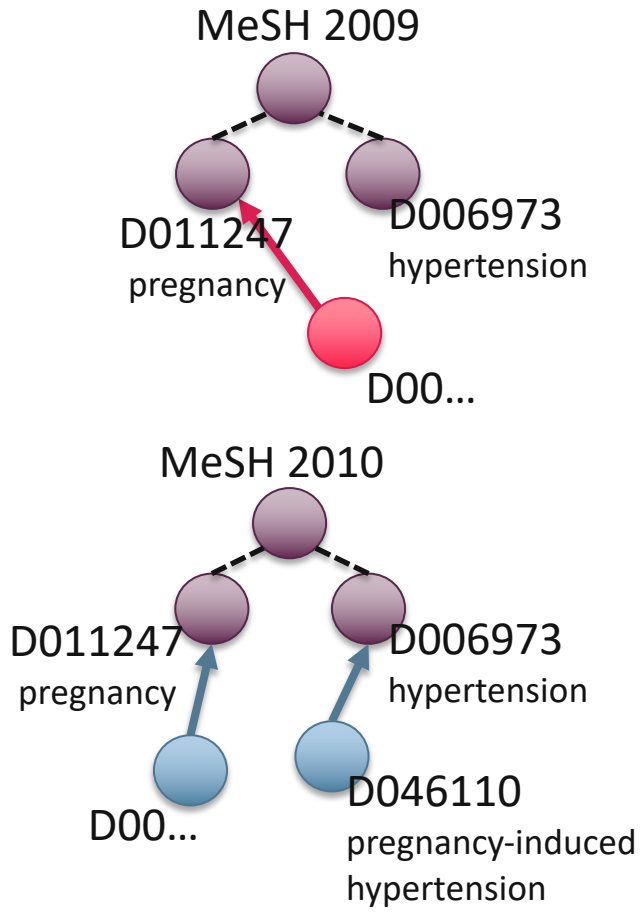
$$V = \left\{ (c, pv, at) \;\middle|\; \begin{array}{l} c \in C, \\ pv = (s, e), s, e \in \mathbb{N}, 0 \leq s \leq e \\ at \in A \end{array} \right\}$$

$$E = \left\{ (u, v, rel) \;\middle|\; \begin{array}{l} u, v \in V, \\ rel \in \{hierarchical, evolutionary\} \end{array} \right\}$$

KNOWLEDGE GRAPH CONSTRUCTION

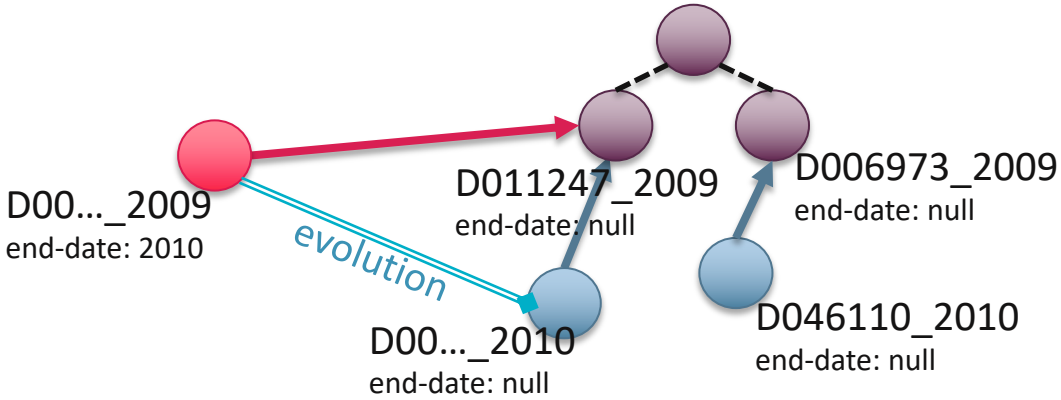
From a pragmatic point of view

1. ”Cleaning” of the initial version
   A. Identification of concepts
   B. Construction of the hierarchy
   C. Definition of validity start date
2. For each new version:
   A. Identification of concepts that are added, modified, deleted using DynDiff
   B. Update of the graph including:
      i. The identified concepts
      ii. Validity date (start and end date)



OUTPUT OF THE ALGORITHM

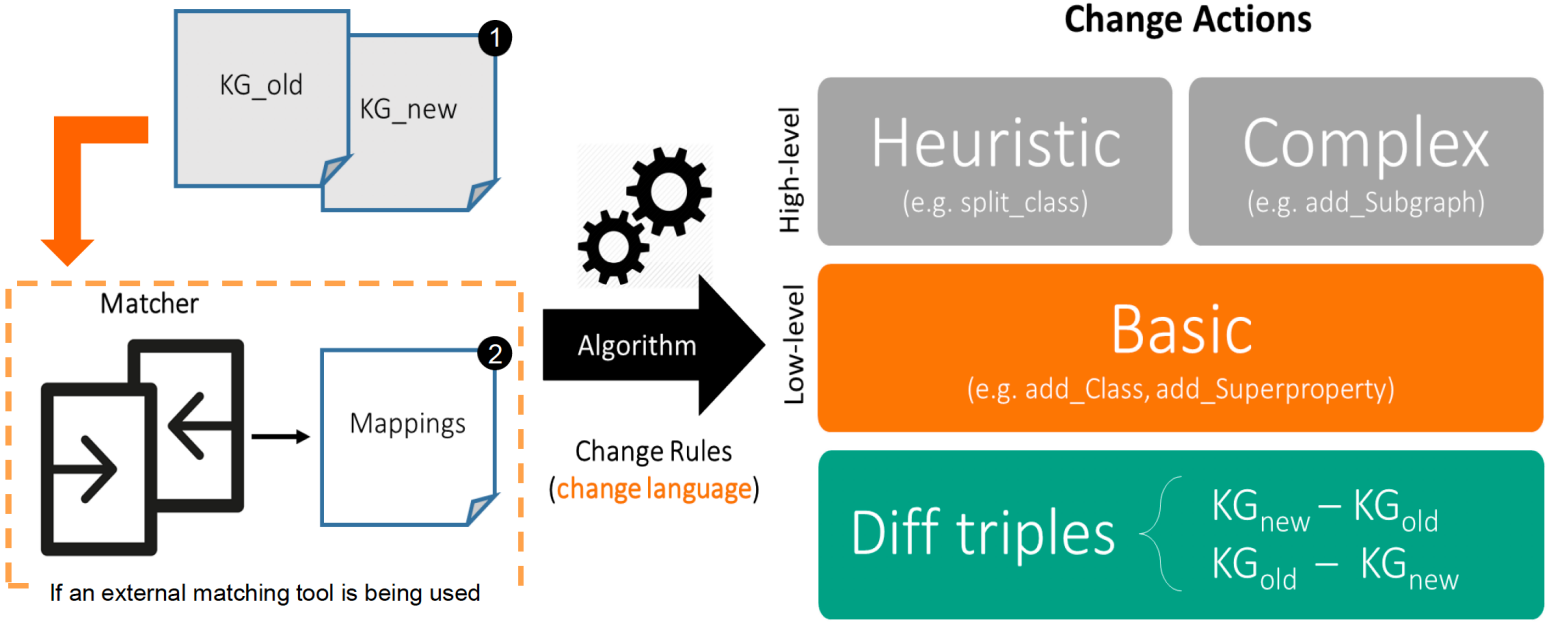
Historical Knowledge Graph



DYNDIFF

A tool for efficient ontology/KG diff computation

LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY | LIST



DYNDIFF

Evaluation

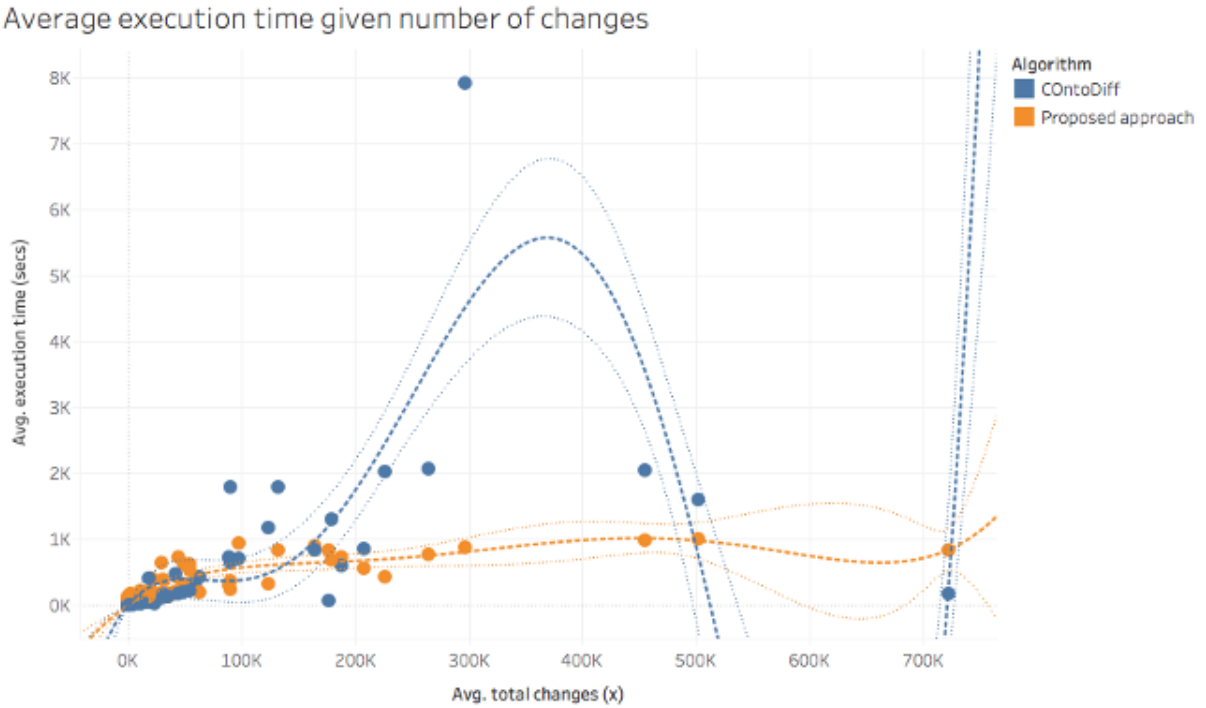
LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY LIST

Acronym	Versions (#)	Average number of					
		C	P	I	R	A	
CIDO	01262020 - 06142020 (14)	3,599	165	294	15,368	25,407	Matcher testing
ICD9CM	2005AA – 2016AA (12)	21,765	2	0	21,765	54,572	
MESH	2005AA – 2016AA (12)	25,820	2	0	34,899	215,971	→ Middle-sized
GO	01072017 - 21072020 (4)	52,574	54	0	492,651	395,285	→ Parsed from OBO
NCIT	2005AA – 2016AA (12)	70,396	2	0	79,525	179,983	Large-sized
IOBC	1.0.0 - 1.4.0 (7)	115,114	84	59,346	707,970	904,443	
SNOMEDCT	2005AA - 2016AA (12)	309,572	2	0	15,368	25,407	

Average number of concepts **C**, properties **P**, instances **I**, relationships **R** and attributes **A**

DYNDIFF

Evaluation (2)

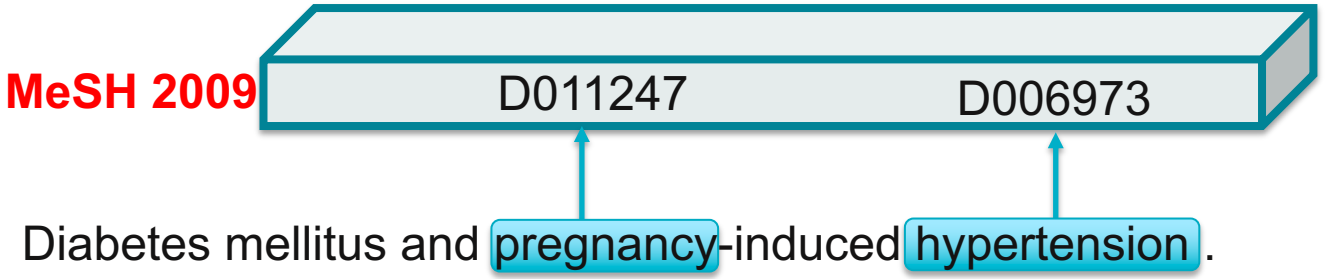


Model formula:	Algorithm*(x^5 + x^4 + x^3 + x^2 + x)
R-Squared:	0.740068
p-value (significance):	< 0.0001

UTILISATION OF THE HKG

Maintenance of semantic annotations

- A semantic annotation is the association of a concept from an ontology with a piece of digital information (e.g. text, image, video …)



- Use of HKG for detecting and updating outdated semantic annotations
- Corpus of 500 annotations produced at 2 different moments in time (2009 et 2016) with NCIt, ICD-9-CM, MeSH and SNOMED CT validated by domain experts
- Comparison of the maintenance task using our HKG with existing method using background knowledge given by BioPortal

EVALUATION

For detecting invalid semantic annotations

LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY | LIST

Method	ICD-9-CM			MeSH		
	P	R	F1	P	R	F1
BK	1	0.129	0.229	1	0.050	0.094
KG	1	0.500	0.667	0.974	0.306	0.465

Method	NCIt			SNOMED CT		
	P	R	F1	P	R	F1
BK	1	0.115	0.207	1	0.625	0.769
KG	0.975	0.750	0.848	0.917	0.668	0.786

EVALUATION

Migration of invalid semantic annotations

LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY | LIST

Method	ICD-9-CM	MeSH	NCIt	SNOMED CT
	AUC	AUC	AUC	AUC
BK	0.597	0.545	0.663	0.75
KG	0.593	0.545	0.615	0.74

- The Area Under the Curve (AUC) denotes the probability to take the good decision for a correct migration independently of the considered invalid annotation

CONCLUSION

… and perspectives

- Proposal for the construction and exploitation of a KG allowing the management of the evolution of the concepts of an ontology over time

- Evaluation of the KG for the maintenance of semantic annotations directly impacted by the evolution of the underlying ontology

- Limitations:
  - HKG deals with concepts only
  - The relationship “evolveTo” is too generic

- Exploitation of the HKG for other use-cases:
  - Information retrieval

Cardoso, S. D., Da Silveira, M., & Pruski, C. (2020). Construction and exploitation of an historical knowledge graph to deal with the evolution of ontologies. *Knowledge-Based Systems*, *194*, 105508.

Thanks!

Cedric.Pruski@list.lu