A First Step towards Extending the Materials Design Ontology

Mina Abd Nikooie Pour, Huanyu Li, Rickard Armiento, and Patrick Lambrix

7 June 2021
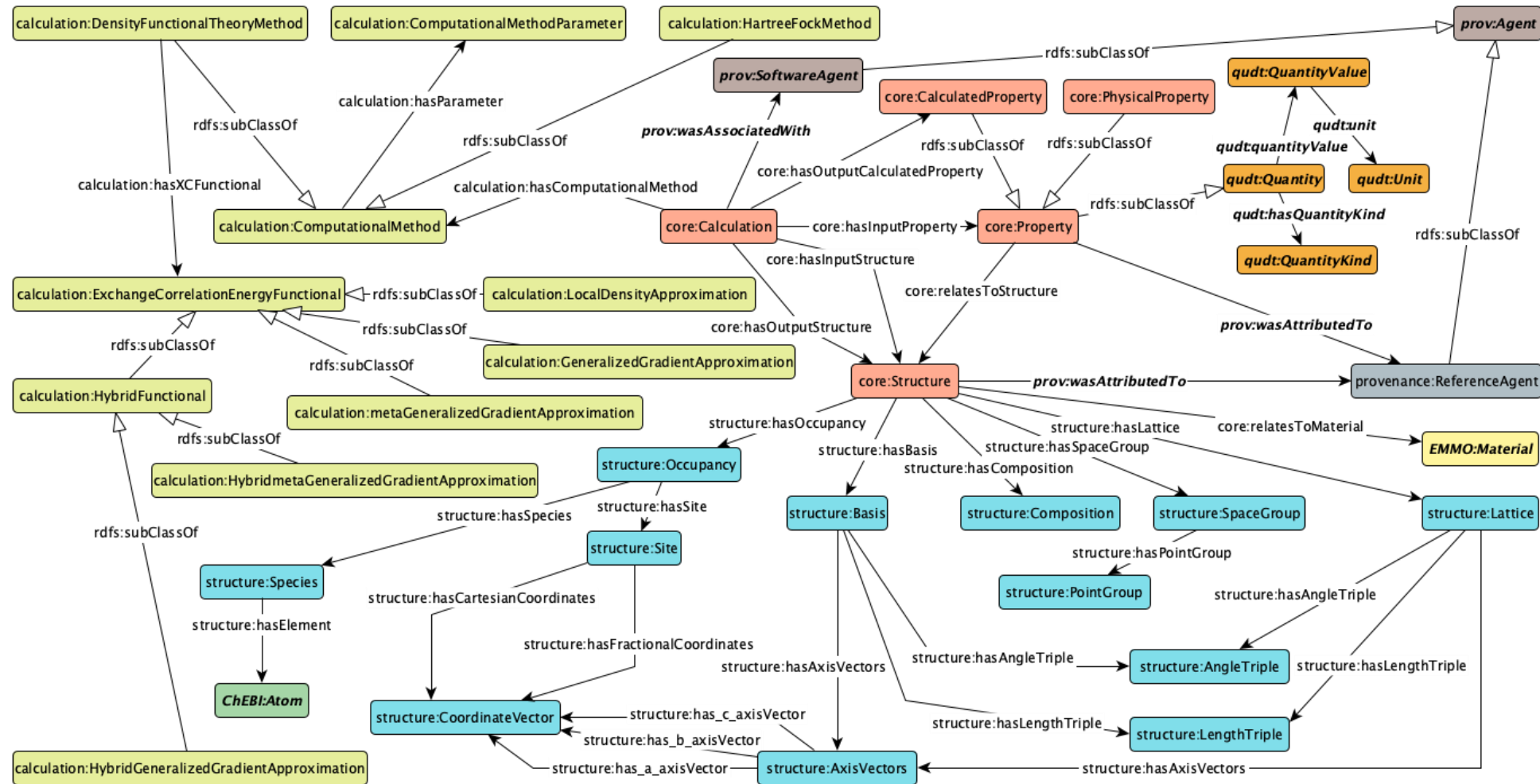DORIC-MM 2021

LINKÖPINGS
UNIVERSITET

Outline

- Introduction

- The Materials Design Ontology (MDO)

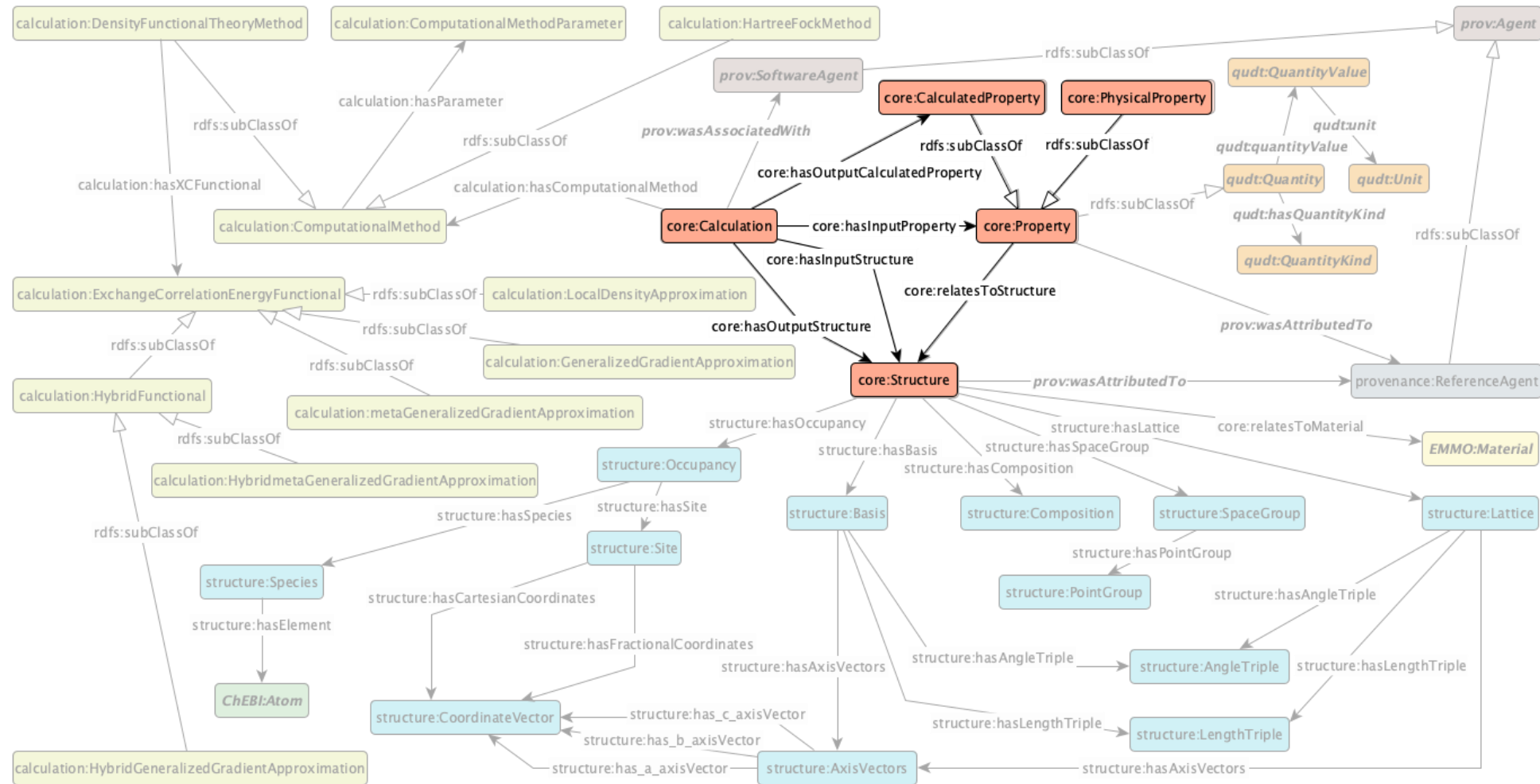- Method for extending ontologies

- Extending the MDO

- Conclusion

LINKÖPINGS UNIVERSITET

Introduction

- Ontologies as a means to make data FAIR

  - **F**indable
  - **A**ccessible
  - **I**nteroperable
  - **R**eusable

- Need for high quality ontologies

LINKÖPINGS UNIVERSITET

The Materials Design Ontology (MDO)

Li, H., Armiento, R., Lambrix, P.: An ontology for the materials design domain. ISWC 2020.

LINKÖPINGS UNIVERSITET

The Materials Design Ontology (MDO)

Li, H., Armiento, R., Lambrix, P.: An ontology for the materials design domain. ISWC 2020.

LINKÖPINGS UNIVERSITET

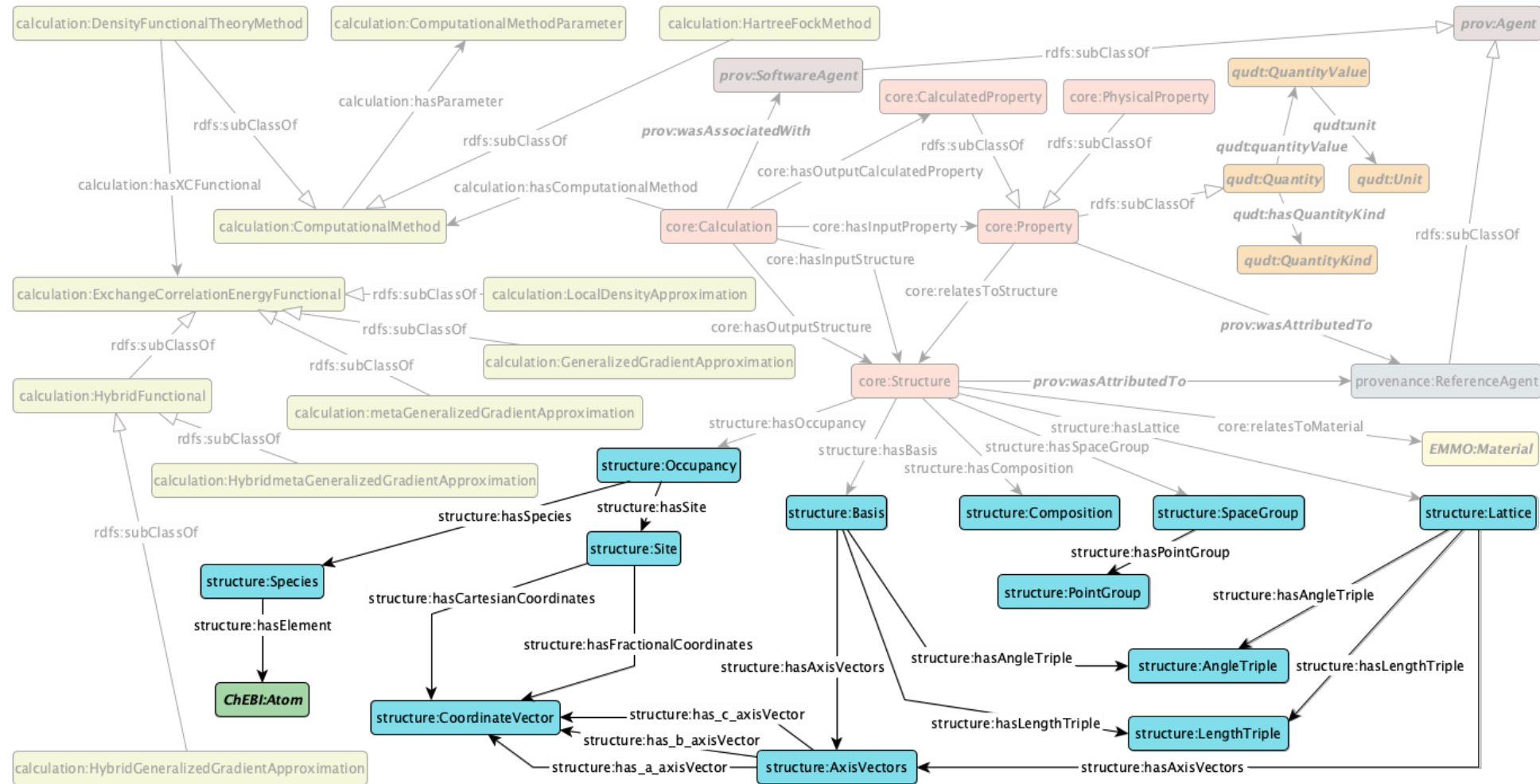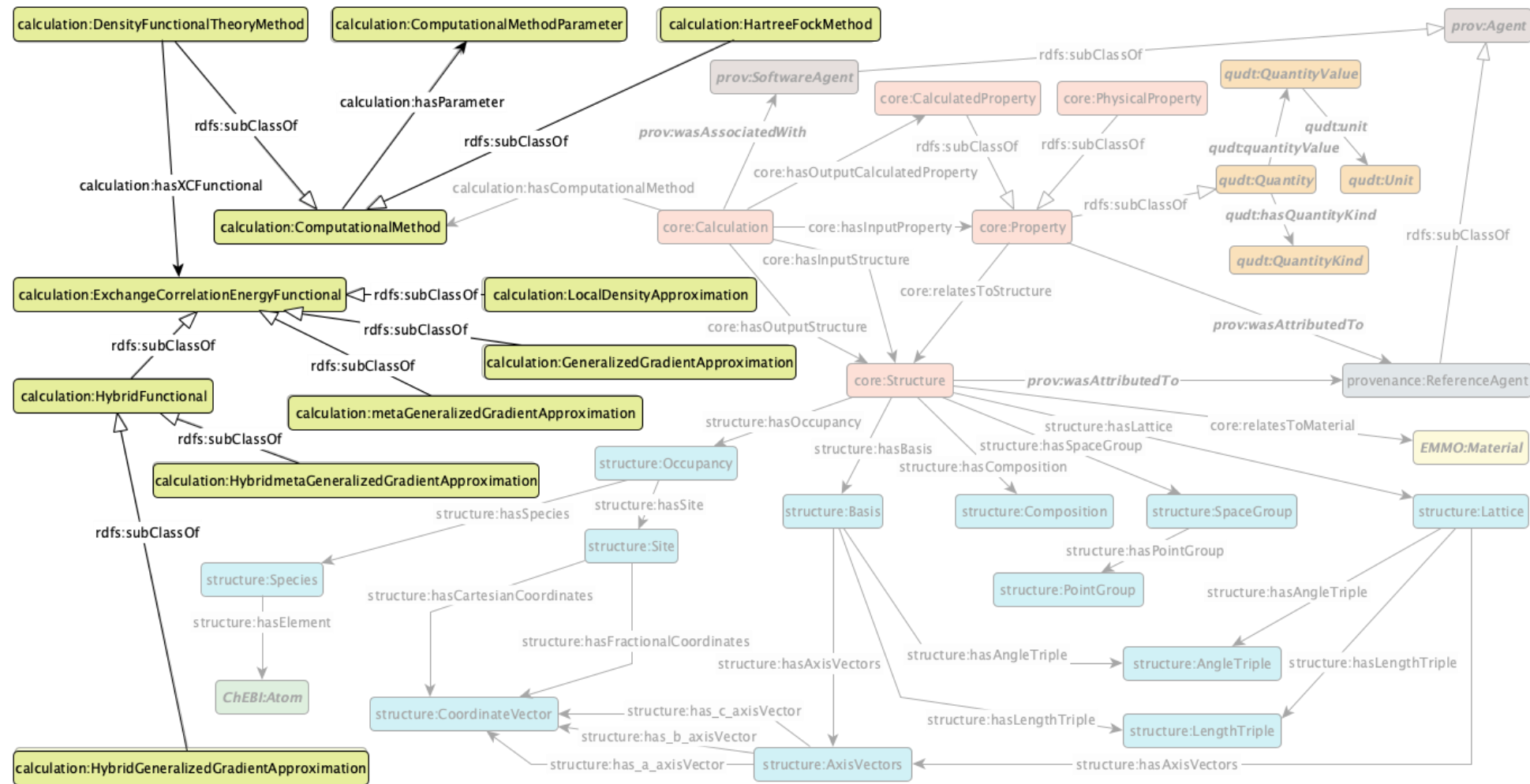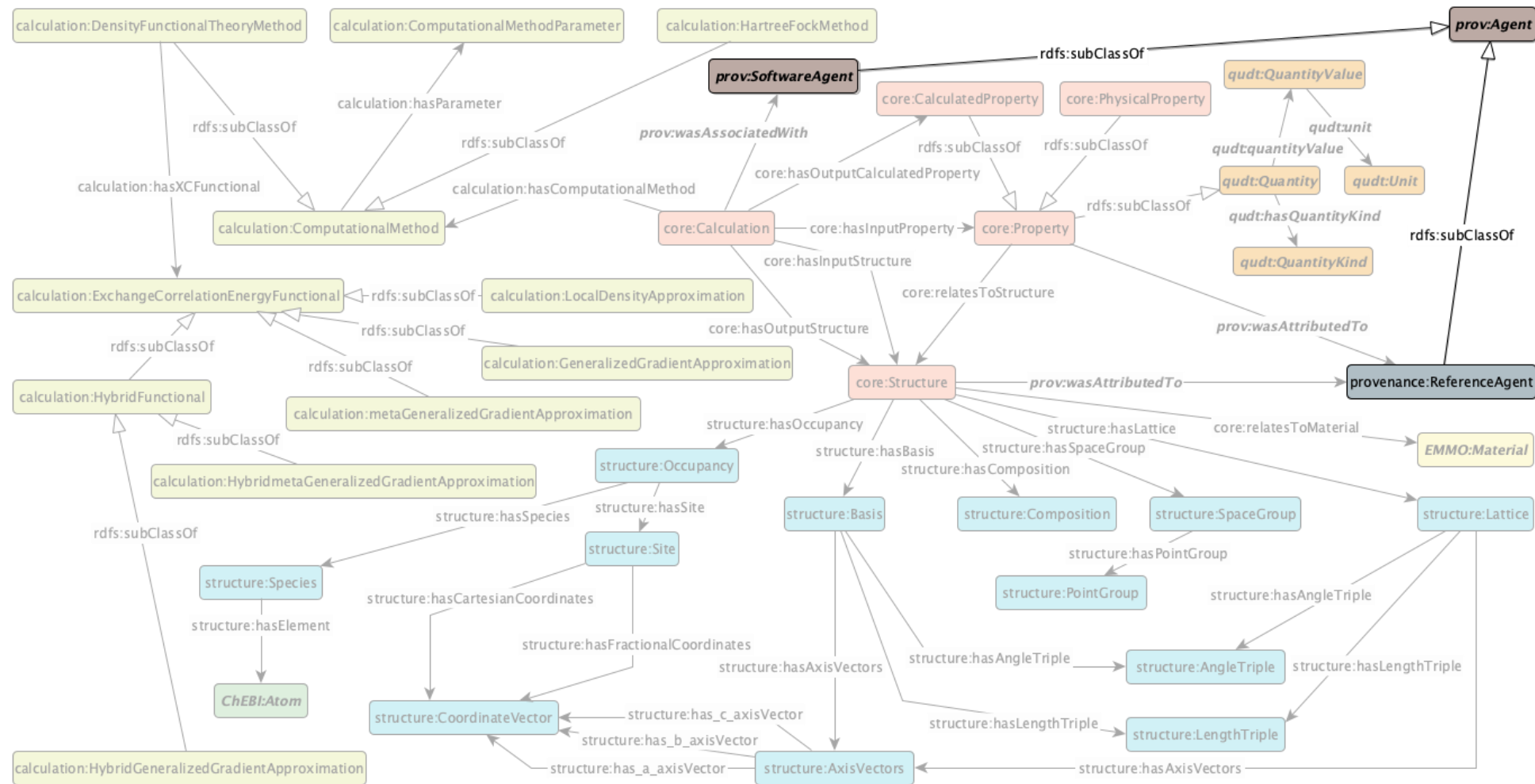# The Materials Design Ontology (MDO)

Li, H., Armiento, R., Lambrix, P.: An ontology for the materials design domain. ISWC 2020.

The Materials Design Ontology (MDO)

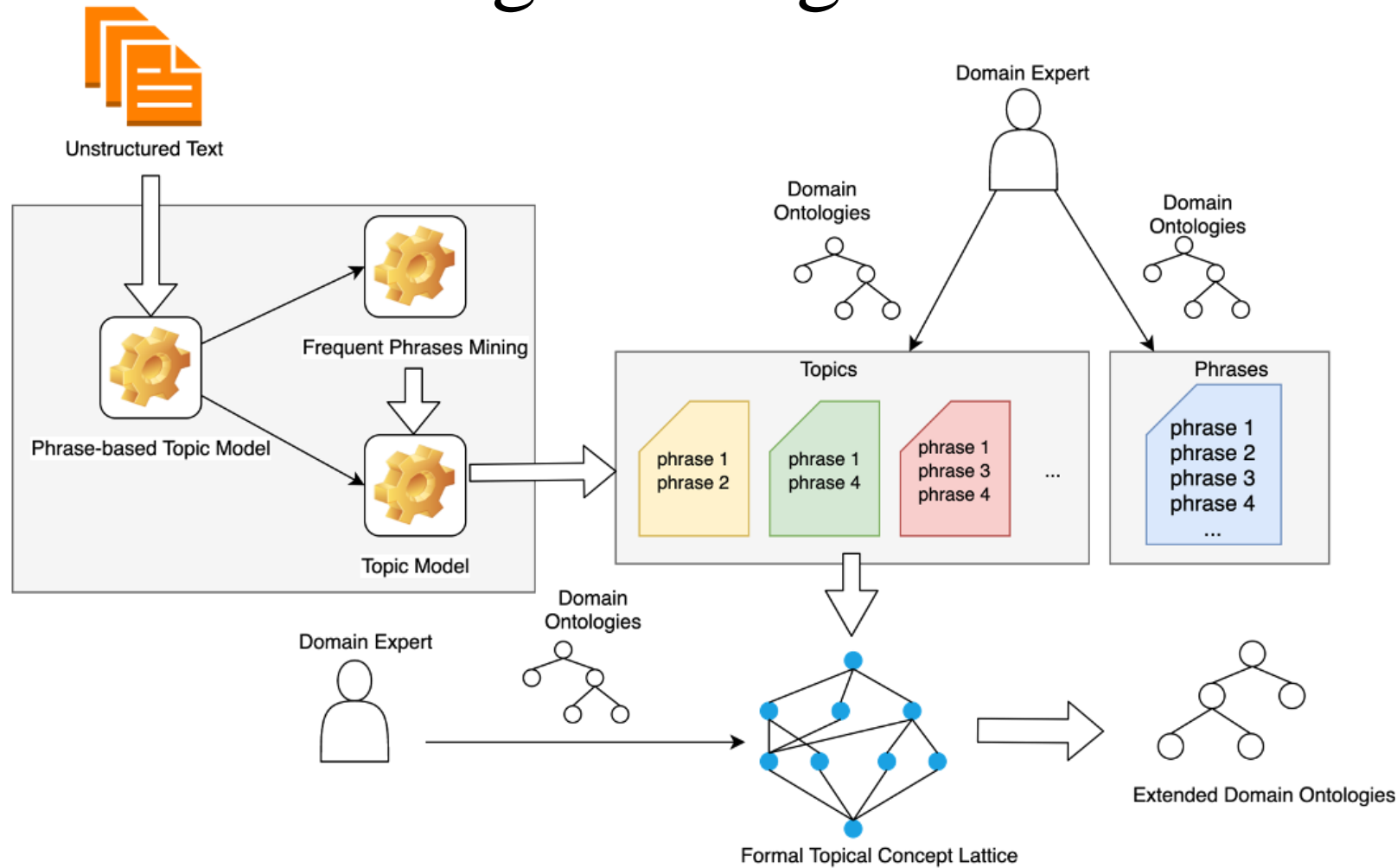Li, H., Armiento, R., Lambrix, P.: An ontology for the materials design domain. ISWC 2020.

The Materials Design Ontology (MDO)

Li, H., Armiento, R., Lambrix, P.: An ontology for the materials design domain. ISWC 2020.

LINKÖPINGS UNIVERSITET
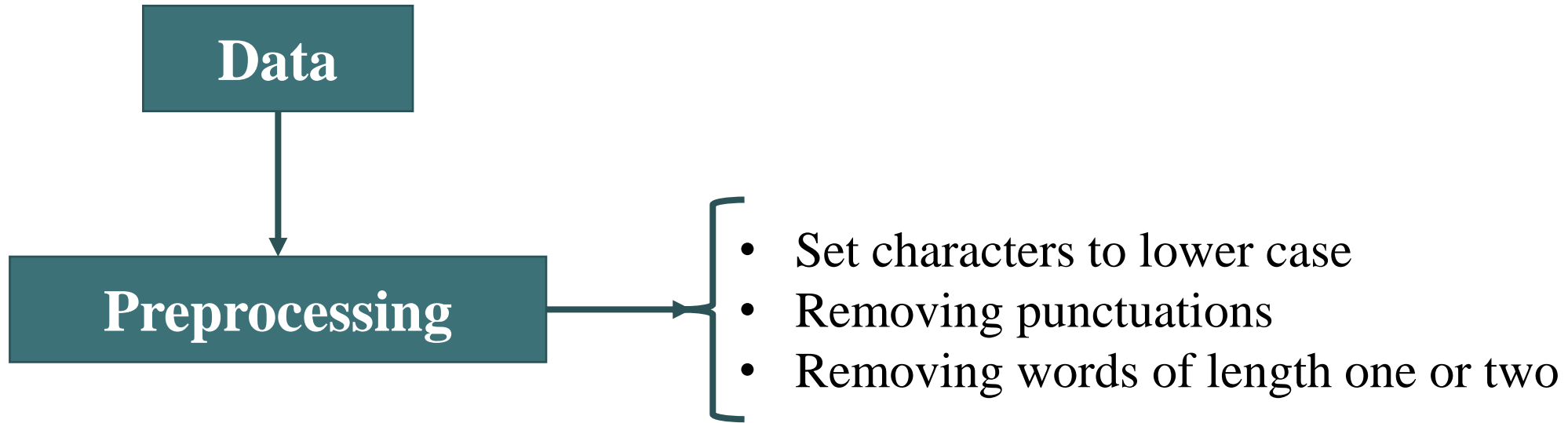
Method for extending ontologies

Li, H., Armiento, R., Lambrix, P.: A method for extending ontologies with application to the materials science domain. Data Science Journal 2019.

Extending the MDO

- Data
  - Two journals in the field of materials design
    - NPJ Computational Materials
    - Computational Materials Science

  - We use the 37 concepts of MDO as search phrases

Name of journal	Number of retrieved articles	Parts of articles to collect
NPJ Computational Materials	**403**	Title + Abstract
Computational Material Science	**8,193**	Title + Abstract

LINKÖPINGS UNIVERSITET

Extending the MDO - Data

Data

↓

Preprocessing →
- Set characters to lower case
- Removing punctuations
- Removing words of length one or two

- After preprocessing:
  - Number of distinct words: 21,548
  - Number of all words: 808,862

LINKÖPINGS UNIVERSITET

Extending the MDO – Data (cont.)

- The distribution of word frequency after preprocessing

Frequency	Percentage of words
less than 10	72.27
10-30	13.25
31-100	7.76
101-500	5.25
501-1000	0.83
1001-2000	0.44
2001-3000	0.12
More than 3000	0.08

"based"	"study"	"electronic"
"properties"	"structure"	"model"
"method"	"temperature"	"molecular"
"calculations"	"density"	"simulations"
"phase"	"results"	"surface"
"materials"	"energy"	

LINKÖPINGS UNIVERSITET

Extending the MDO - Frequent phrases

- **Frequent phrases**:
  - Phrases occur at least min_support times
    - min_support = minimum support threshold

  - New defined threshold:
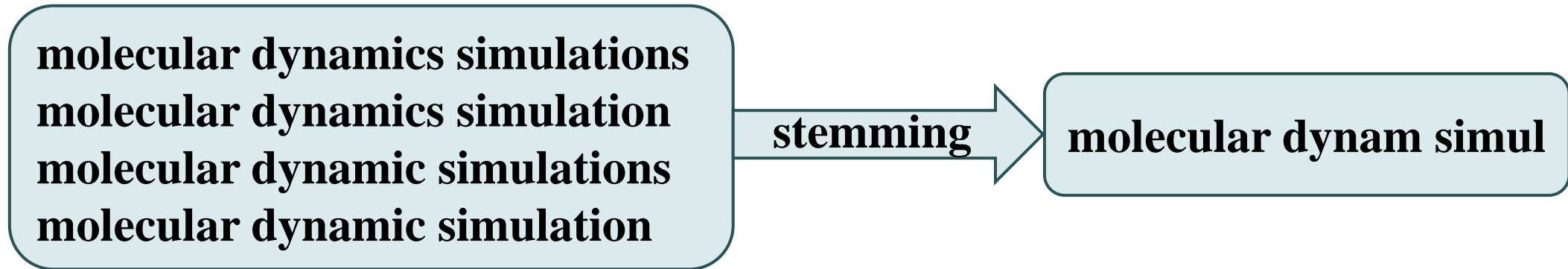   - max_support_word = maximum support threshold for words

- **New ToPMine:**
  - ToPMine algorithm with adding max_support_word as well as the preprocessing step
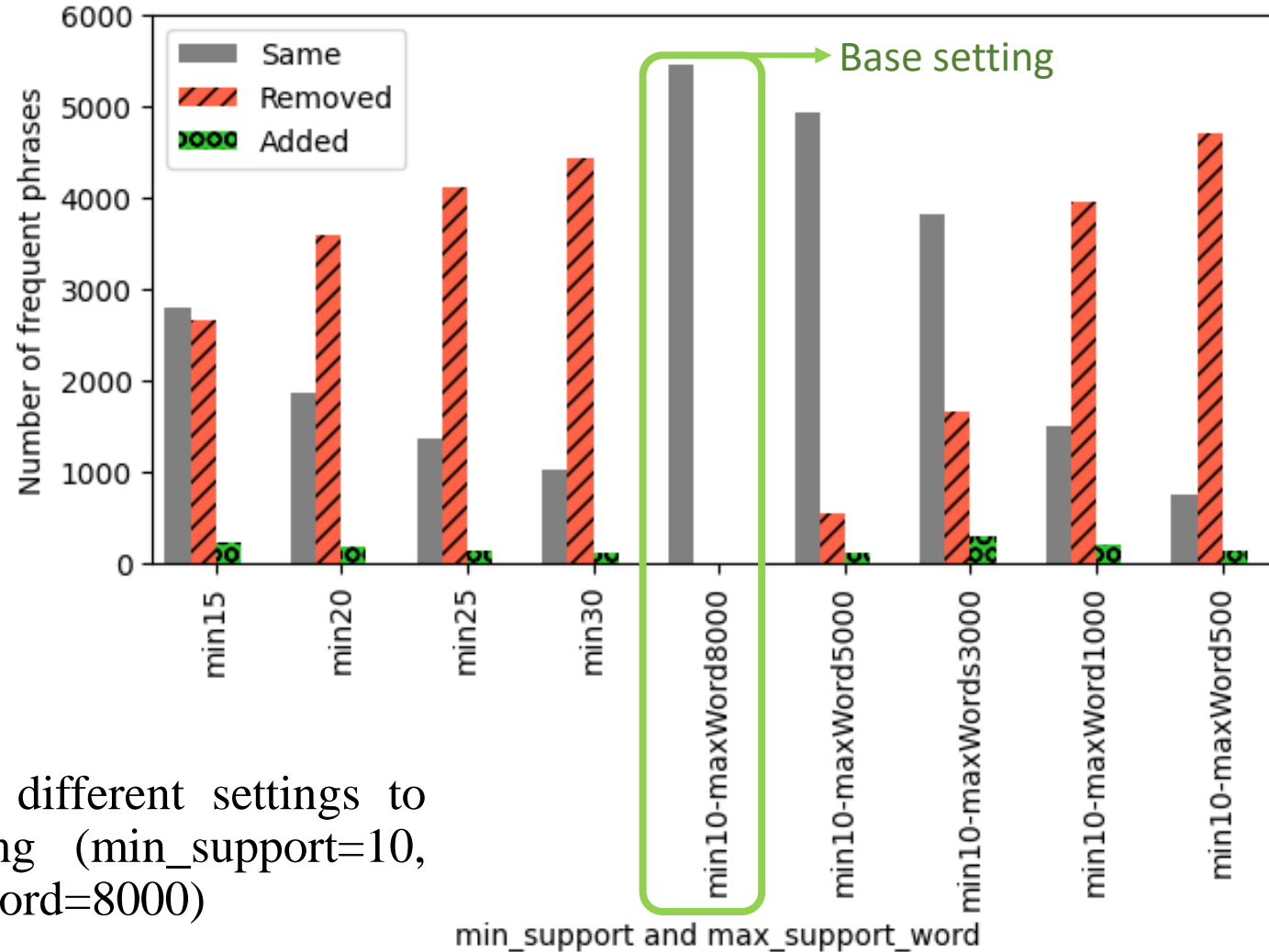
El-Kishky, A., Song, Y., Wang, C., Voss, C.R., Han, J.: Scalable topical phrase mining from text corpora. Proceedings of the VLDB Endowment 2014.

Extending the MDO - Frequent phrases (cont.)

- Using of stemming in mining frequent phrases
  - Removing redundant phrases
  - Reducing the work of the domain expert

molecular dynamics simulations
molecular dynamics simulation
molecular dynamic simulations
molecular dynamic simulation

stemming →

molecular dynam simul

LINKÖPINGS UNIVERSITET

Extending the MDO - Frequent phrases (cont.)

- Comparison of different settings to the base setting (min_support=10, max_support_word=8000)
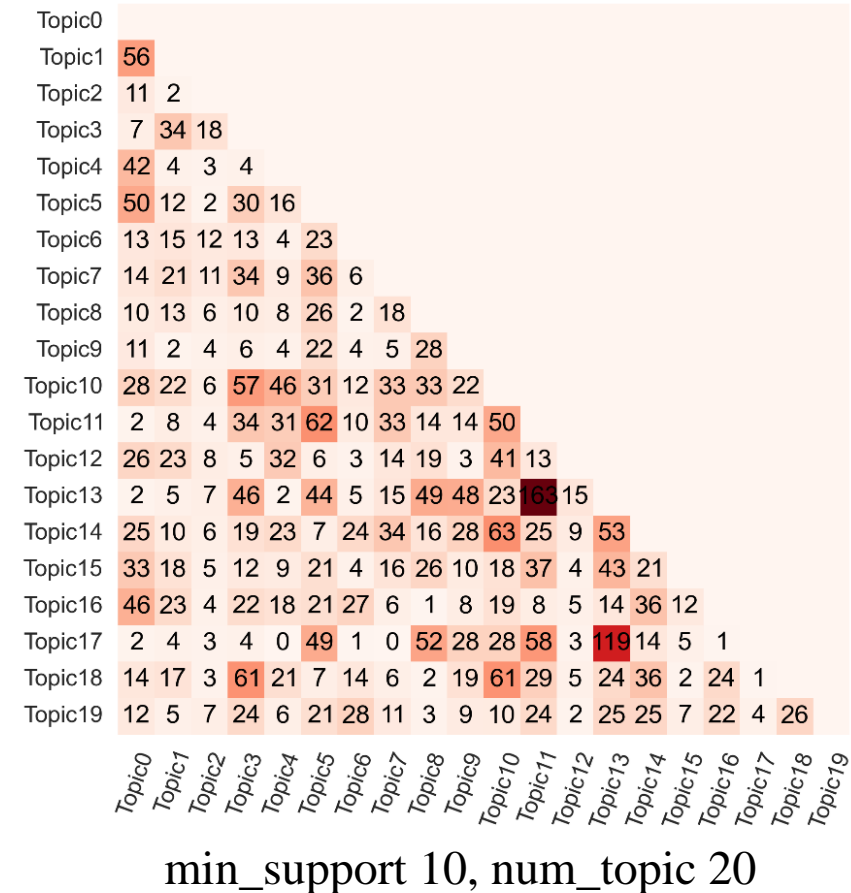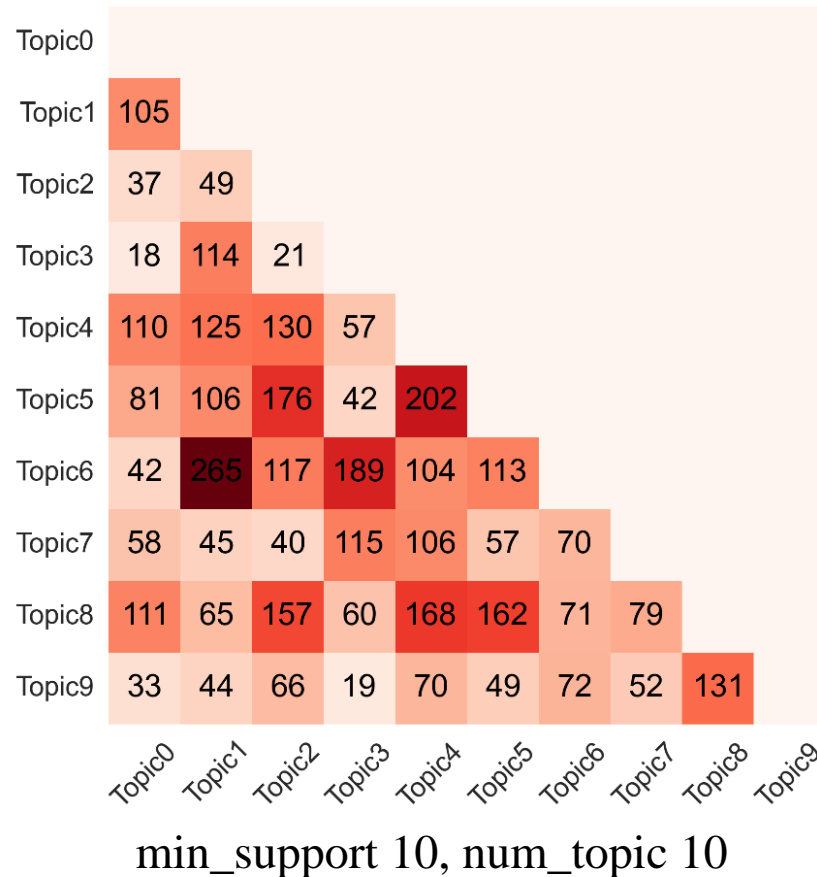
Extending the MDO - Frequent phrases (cont.)

- Suggestion of 88 candidate concepts by the domain expert
  - min_support = 30, max_support_word = 500
  - Based on 81 out of 131 frequent phrases generated by the experiment

Stacking Fault	Stone-wales Defect	Cement Paste
Van der Waals Force	Covalent Bond	Perdew-Burke-Ernzerhof (PBE) Exchange-Correlation Functional
Functionally Graded Material	Symmetric Tilt Grain Boundary Structure	Fatigue Limit
Linearized Augmented Plane Wave Method	Asymmetric Tilt Grain Boundary Structure	Edurance Limit
Face Centered Cubic	Rock Salt Structure	Porous Media
Boron Nitride	Rock Salt	Microstructural Features
Nearest Neighbor	Projector Augmented Wave Method	Hall-Petch Relation
Body Centered Cubic	Iron	Conduction Band
Coarse Grained Model	Cahn–Hilliard Equation	Slip Plane
Fiber Reinforced	Cauchy-Born Rule	Vapor Deposition
Zinc Blende	Domain Wall	Spinodal Decomposition

LINKÖPINGS UNIVERSITET

Extending the MDO - Topics

- Each topic contains a set of phrases that do not have to be disjoint

min_support 10, num_topic 10

min_support 10, num_topic 20

Extending the MDO – Topics (cont.)

- Part of topic labelling based on domain expert validation of frequent phrases with min_support 30 and max_support_word 500

Topic NO.	Topic Labels	Representative Phrases
1	Hardness-related Materials Concepts	Quasi-harmonic Debye Model
		Quasi-harmonic Model
		Rock Salt
		Sound Velocity
		Zinc Blende
2	Materials Strength-related Concepts	Stacking Fault
		Van der Waals Force
		Tension Compression
		Uniaxial Tension
		Symmetric Tilt Grain Boundary Structure

LINKÖPINGS UNIVERSITET

Conclusion

- We started our work on extending MDO using a phrased-based topic model.

- We investigated the influence of different settings on the number of frequent phrases that are generated.

- **Future work**
  - Continuing to validate the results of the different variants and settings

  - Implementing a system to facilitate phrase validation

LINKÖPINGS UNIVERSITET