Extending Ontologies in the Nanotechnology Domain using Topic Models and Formal Topical Concept Analysis on Unstructured Text

Huanyu Li, Rickard Armiento, Patrick Lambrix

The Swedish e-Science Research Centre & Linköping University

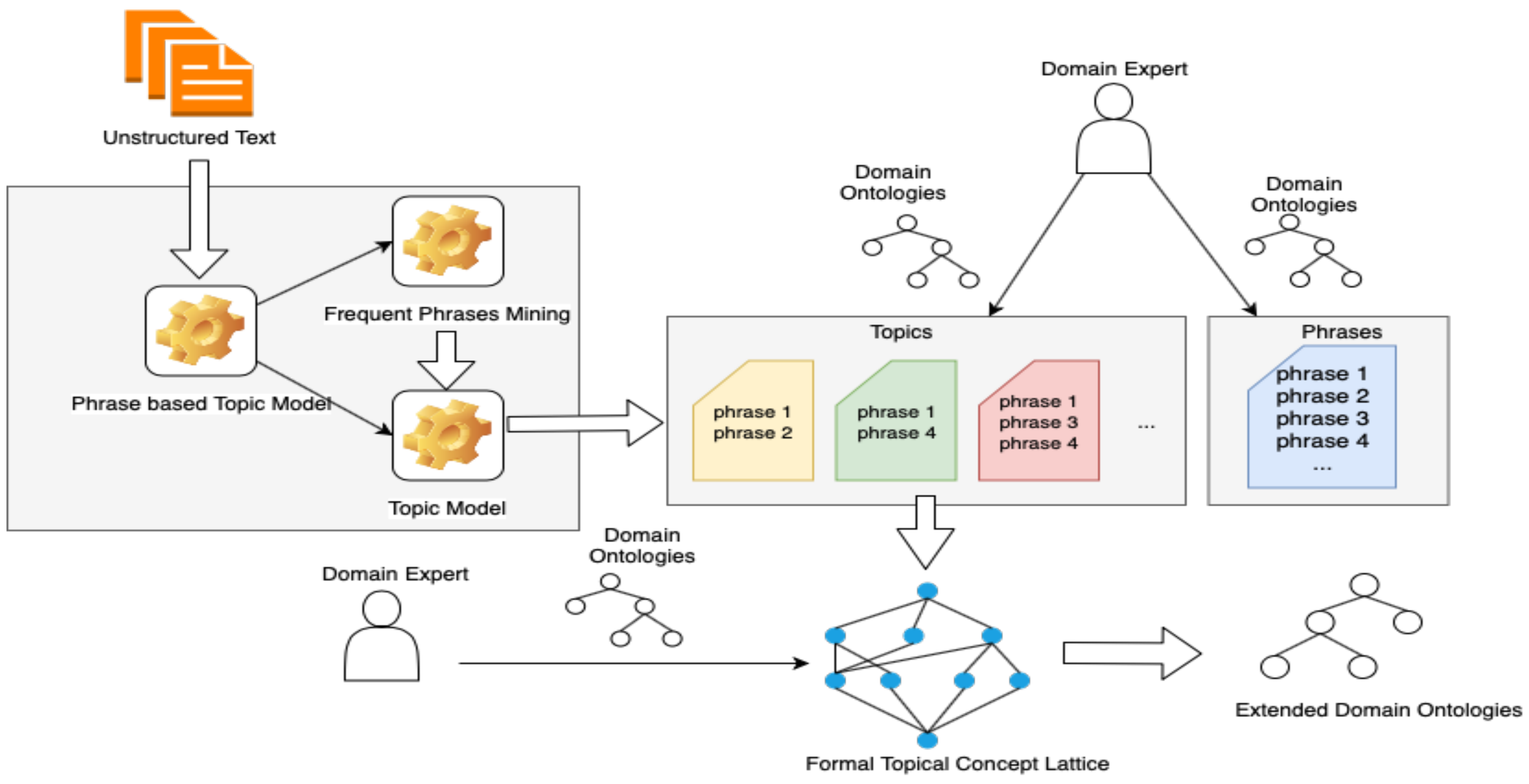
Linköping, Sweden

Motivation

Since the 2000s, materials science has shifted towards the data-driven science paradigm. As more and more data-driven techniques become available, also big data challenges and challenges regarding data integration and reuse appear [1]. Ontologies are seen as a technology that can alleviate these problems. However, there are not so many ontologies yet in this domain and even more mature ontologies such as NanoParticle Ontology (NPO) and eNanoMapper are not complete. Extending such domain ontologies would enhance the quality of semantically-enabled applications that use these ontologies.

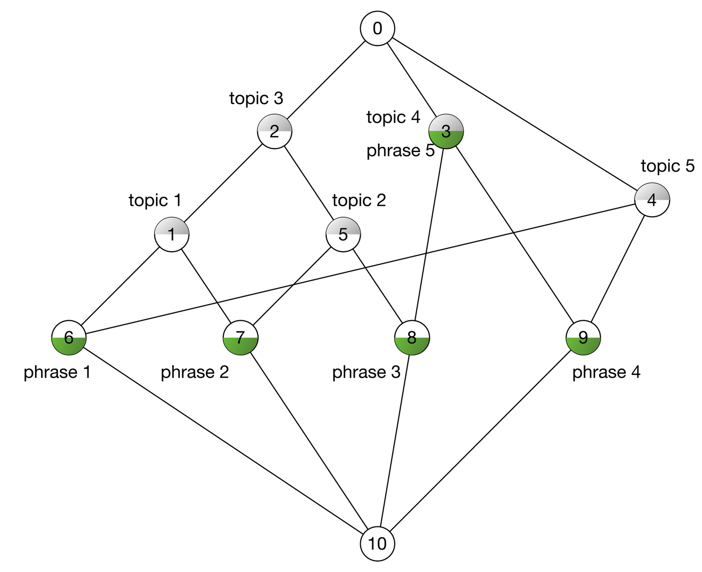
Approach

- ❑ Given a corpus of documents, create a phrase-based topic model.
- ❑ Apply formal topical concept analysis on the result of the topic model.
- ❑ A domain expert interprets all phrases, and topics, and the formal concept analysis-based lattice.
- ❑ The outcome can be divided into
  - ➢ Existing knowledge (*EXIST(-m)*)
  - ➢ New knowledge (*ADD(-m)*)
  - ➢ Too specific concept (*Q*)
  - ➢ Too general concept (*No-g*)
  - ➢ No meaningful finding (*No*)
  - ✻ *-m means modified*



	topic 1	topic 2	topic 3	topic 4	topic 5
phrase 1	√		√		√
phrase 2	√	√	√		
phrase 3		√	√	√	
phrase 4				√	√
phrase 5				√	

Phrase occurrences in topics



Formal Topical Concept Lattice

ID	Formal Topical Concept	ID	Formal Topical Concept
0	({phrase 1, phrase 2, phrase 3, phrase 4, phrase 5}, ∅)	6	({phrase 1}, {topic 1, topic 3, topic 5})
1	({phrase 1, phrase 2}, {topic 1, topic 3})	7	({phrase 2}, {topic 1, topic 2, topic 3})
2	({phrase 1, phrase 2, phrase 3}, {topic 3})	8	({phrase 3}, {topic 2, topic 3, topic 4})
3	({phrase 3, phrase 4, phrase 5}, {topic 4})	9	({phrase 4}, {topic 4, topic 5})
4	({phrase 1, phrase 4}, {topic 5})	10	(∅, {topic 1, topic 2, topic 3, topic 4, topic 5})
5	({phrase 2, phrase 3}, {topic 2, topic 3})		

Formal Topical Concepts

Results

- ❑ Examples:
  - ❖ *'carbon nanotube' exists in NPO. (EXIST)*
  - ❖ *For 'force microscopy', 'atomic force microscopy' exists in NPO. (EXIST-m)*
  - ❖ *'CdTe nanoparticle' should be added into NPO. (ADD)*
  - ❖ *For 'reverse micelle', 'reverse micelle-type quantum dot' should be added into NPO. (ADD-m)*
  - ❖ *Labelled topic 'optical properties of copper nanoparticles and CdTe nanoparticles'. (Q)*
  - ❖ *'ground state' is a general concept in materials science. (No-g)*
  - ❖ *'single electron' does not contribute to NPO. (No)*

	label	NPO	eNanoMapper
# of general concepts	No-g	14	12
# of existing concepts	EXIST(-m)	46	52
# of new concepts	ADD(-m)	35	32
# of new axioms	ADD(-m)	42	37
# of queries	Q	49	49
# of influenced concepts	-	72	37

Discussion

The above work leads to the confirmation of ontological concepts or the addition of new concepts and axioms. The domain expert noted that the topical concept lattice helps in refining topics. More support may be needed to help the domain expert in deciding on the level of granularity for concepts. More details can be found in [2].

[1] Lambrix, P., Armiento, R., Delin, A., Li, H., 2018. Big Semantic Data Processing in the Materials Design Domain. Encyclopedia of Big Data Technologies, DOI: 10.1007/978-3-319-63962-8_293-1

[2] Li, H., Armiento, R. and Lambrix, P., 2019. A Method for Extending Ontologies with Application to the Materials Science Domain. Data Science Journal, 18(1), p.50. DOI: 10.5334/dsj-2019-050

LINKÖPING UNIVERSITY

SeRC Swedish e-Science Research Centre