

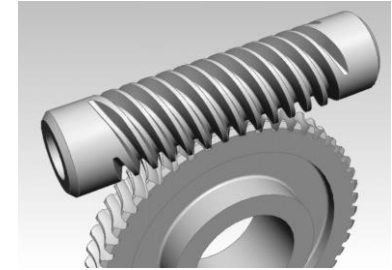
**ADVANCING SCIENCE**

# Deep learning services for patents

Mustafa Sofean and Ahmad Alrifai

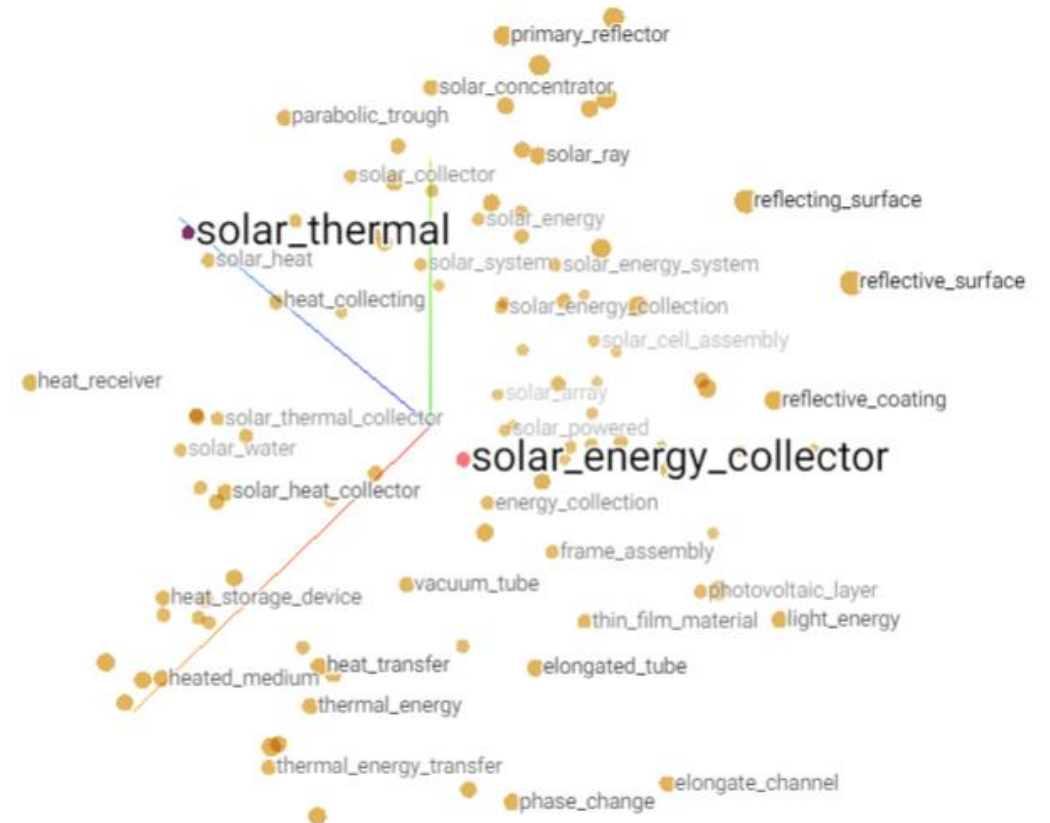
# Motivation

- Technical terms and Multi-Word Units (MWUs)
  - “gear” – “a toothed wheel” – “profiled wheel with teeth”
  - “camera” – “optical device” – “automatic focusing apparatus”
- Patent search is recall based
  - Query expansion using Controlled vocabularies
- phrase embedding: different vectors for different senses
  - Distinguish “*vegetable oil*” and “*organic oil*” from “*vehicle oil*”
- Domain-specific word and phrase embedding model



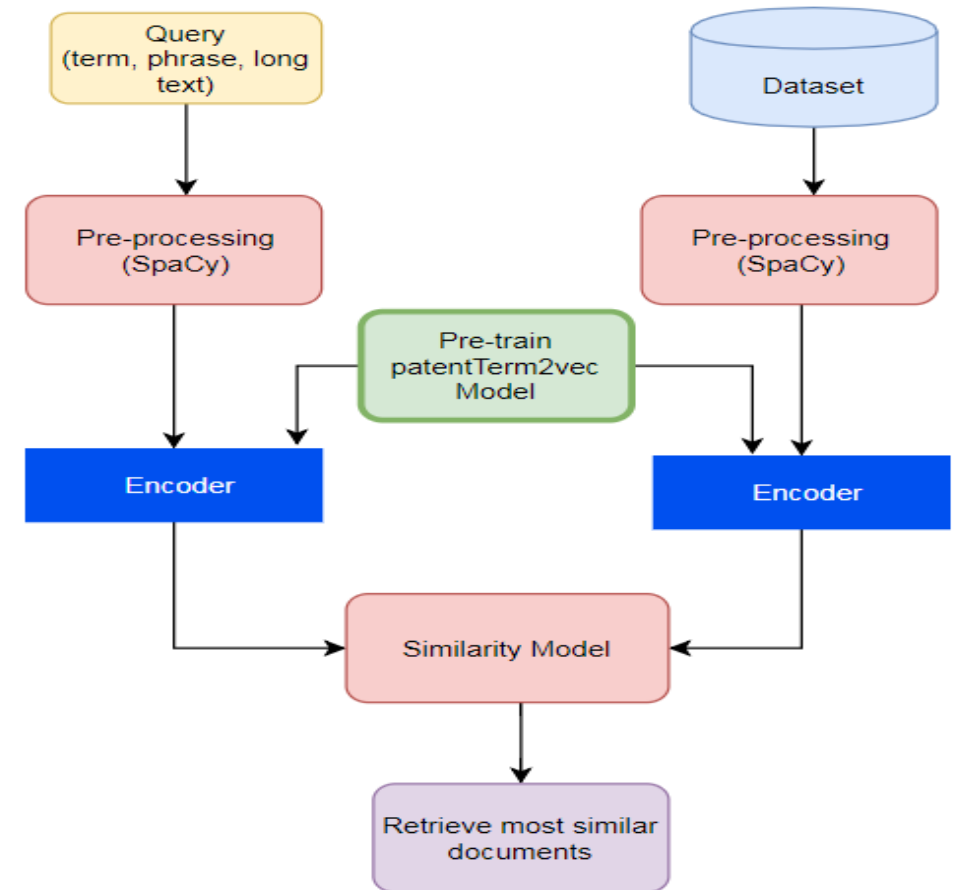
# Approach

- NLP preprocessing pipeline
  - Sentence detection
  - Tokenization, POS-tagging
  - Shallow Syntactic parsing (Chunking)
  - Concatenation: “digital\_rights\_management”
- Train a word2vec model on the preprocessed text
  - 5,700,000 titles and abstracts of English patents
  - Skip-gram model: vector size 100, and window size of 10



# Services and applications

- Similarity between two words/phrases
  - Finding synonyms as a use case
- Patent semantic similarity search
- Evaluation with internal experts
  - To do: Preparing a collecting data sets for evaluation and comparison
- Future step: Enrichment of the patent classification schemes



# THANK YOU!

## Demo time

### Contact

Mustafa Sofean  
Senior Researcher

Mustafa.Sofean@fiz-Karlsruhe.de

Ahmad Alrifai  
Senior Researcher

Ahmad.Alrifai@fiz-Karlsruhe.de

© FIZ Karlsruhe 2019  
Leibniz-Institut für Informationsinfrastruktur GmbH  
[www.fiz-karlsruhe.de](http://www.fiz-karlsruhe.de)



Except where otherwise noted, content is licensed under a Creative Commons Attribution 4.0 International License.

 **FIZ Karlsruhe**  
Leibniz Institute for Information Infrastructure



*Leibniz*  
Leibniz  
Association