Facilitating the definition and improving the quality of domain models using an NLP-based assistant

Lola Burgueño
Open University of Catalonia, Barcelona, Spain

@LolaBurgueno
lburguenoc@uoc.edu

May 17th, 2022
Introduction

• Gmail autocompletion...
Introduction

Domain modeling

Informal descriptions

Structured and unambiguous representation leaving out superfluous details

... using a concrete notation
Motivation

To promote and facilitate the creation and manipulation of domain models

→ Broad variety of
  • languages
  • tools

Still, they are typically created by hand!!!
Data has a better idea
may have
The role of data

Contextual knowledge

General knowledge
Natural Language Processing (NLP)
Our approach

Partial model

[Diagram showing the process]

Domain docs

Text preprocessing algorithm

Domain corpus of text

preprocess

L. Burgueño, R. Clarisó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106
Our approach

NLP components

- Partial model
- Domain docs
- Text preprocessing algorithm
- Domain corpus of text
- NLP method for word embeddings
- NLP models
  - general knowledge
  - contextual knowledge

L. Burgueño, R. Clarisó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106
Our approach

NLP components

Partial model

Domain docs

Text preprocessing algorithm

Domain corpus of text

NLP method for word embeddings

NLP models

- general knowledge
- contextual knowledge

Morphological analysis & lemmatization

L. Burgueño, R. Clarisó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106
Morphological analysis & lemmatization

1. Perform a morphological analysis of each word
   Part-of-Speech (POS) tagging

2. Lemmatization process
   flyers lemmatized as a verb → fly
   flyers lemmatized as noun → flyer

We use WordNet
part of the Python NLTK (Natural Language Toolkit)
Our approach

L. Burgueño, R. Clarisó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106
Our approach

Model Recommendation Engine

- activity: Recomm. Engine
- slice: B.1
- model slices

Partial model

- Text preprocessing algorithm
- Domain corpus of text
- NLP method for word embeddings
- NLP models
  - general knowledge
  - contextual knowledge
- Morphological analysis & lemmatization
- Morphological analysis & lemmatization

Domain docs

- Domain corpus of text
- A.1 preprocess

A.2 train

Slice 1
- Flight
- Pilot

Slice 2
- Flight
- Pilot

Slice 3
- Pilot
- Pilot
- Flight

Slice 4
- Pilot
- Flight

Slice 5
- Pilot
- Pilot
- Flight
Our approach

- **Partial model**
  - **Domain docs**
    - **Text preprocessing algorithm**
      - A.1 preprocess
    - **Domain corpus of text**
      - A.2 train
  - **NLP method for word embeddings**
  - **NLP models**
    - General knowledge
    - Contextual knowledge
  - **Morphological analysis & lemmatization**

- **Model Recommendation Engine**
  - **activity Recomm. Engine**
    - B.1 slice
    - B.2 query
    - B.3 uses
    - B.4 uses
    - B.5 uses

- **Slice 1**: Flight → [...]
- **Slice 2**: Pilot → [...]
- **Slice 3**: Pilot → [...]
- **Slice 4**: Pilot, Flight → [...]
- **Slice 5**: Pilot, Flight → [...]

14
Our approach

L. Burgueño, R. Clarisó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106
Our approach

Slice 1: Flight → [plane, airline, airplane, jet, flying, fly]

L. Burgueño, R. Clarisó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106
Our approach

How are the potential model elements that will be recommended built?

L. Burgueño, R. Clarisó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106
Our approach: Build

Different slices have different purposes

The model is sliced according to these patterns:

- one slice that contains all the classes in the model after removing their features (attributes and relationships)
  → aims to suggest new classes
  → remove verbs, adjectives and plural nouns
  → the remaining words are candidates to become new classes

- one slice for each class C in the model (keeping its attributes and dangling relationships)
  → aims to suggest new attributes

- one slice for each pair of classes (keeping its attributes and dangling relationships)
  → aims to suggest new relationships
Our approach: Build

Different slices have different purposes

The model is sliced according to these patterns:

- one slice that contains all the classes in the model after removing their features (attributes and relationships)
  → aims to suggest new classes
  → remove verbs, adjectives and plural nouns
  → the remaining words are candidates to become new classes

- one slice for each class C in the model (keeping its attributes and dangling relationships)
  → aims to suggest new attributes

- one slice for each pair of classes (keeping its attributes and dangling relationships)
  → aims to suggest new relationships
Our approach

Add class named Plane
Add class named Airline
Add class named Airplane...

L. Burgueño, R. Clarisó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106
Our approach

Add class named Plane
Add class named Airline
Add class named Airplane

L. Burgueño, R. Clarisó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106
Our approach

L. Burgueño, R. Clarísó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106
Our approach

L. Burgueño, R. Clarisó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106
Conclusions & Future Work

• First step towards a more general **modeling assistant** that effectively helps specify **better models faster**

Future!

• Integrate other types of information (e.g., past models, ontologies,...)
• Work on the usability aspect & empirical study
• Refine the techniques presented in this paper
• Application on other types of models and modeling languages
• Exploitation of other NLP models
Facilitating the definition and improving the quality of domain models using an NLP-based assistant

Lola Burgueño
Open University of Catalonia, Barcelona, Spain

May 17th, 2022