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

geodes Research Day AND Symposium

Lola Burgueño

Open University of Catalonia, Barcelona, Spain

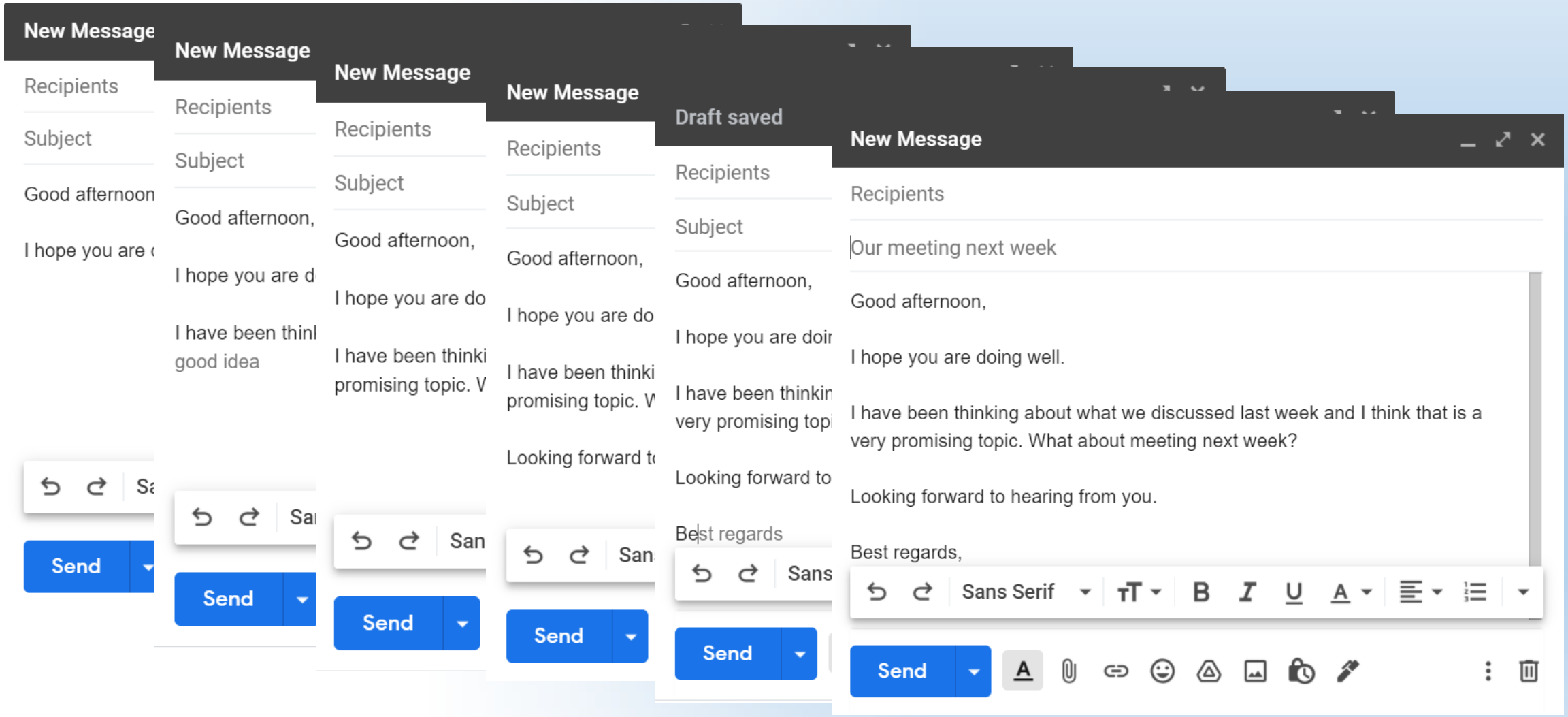
@LolaBurgueno

lburguenoc@uoc.edu

May 17th, 2022

Introduction

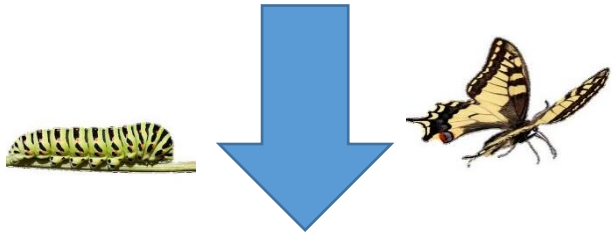
- Gmail autocompletion…



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!!!

DSL ER

UML

emf
ECLIPSE MODELING FRAMEWORK


Data has a better idea
may have

The role of data

WIKIPEDIA
The Free Encyclopedia

English
6 299 000+ articles

Español
1 684 000+ artículos

日本語
1 268 000+ 記事

Deutsch
2 576 000+ Artikel

Русский
1 724 000+ статей

Français
2 329 000+ articles

Italiano
1 693 000+ voci

中文
1 197 000+

Português
1 066 000+ artigos

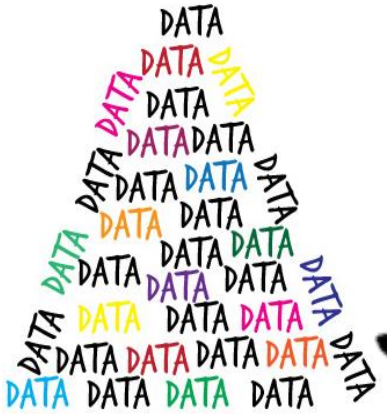
Polski
1 473 000+ haseł

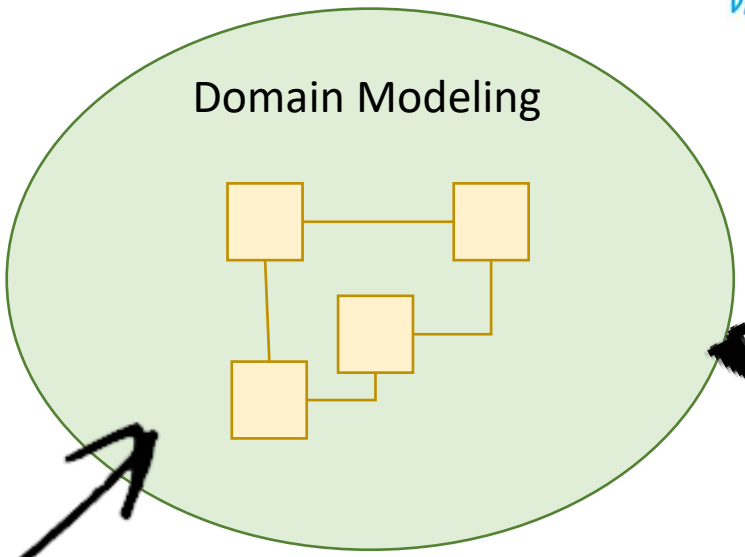
reddit

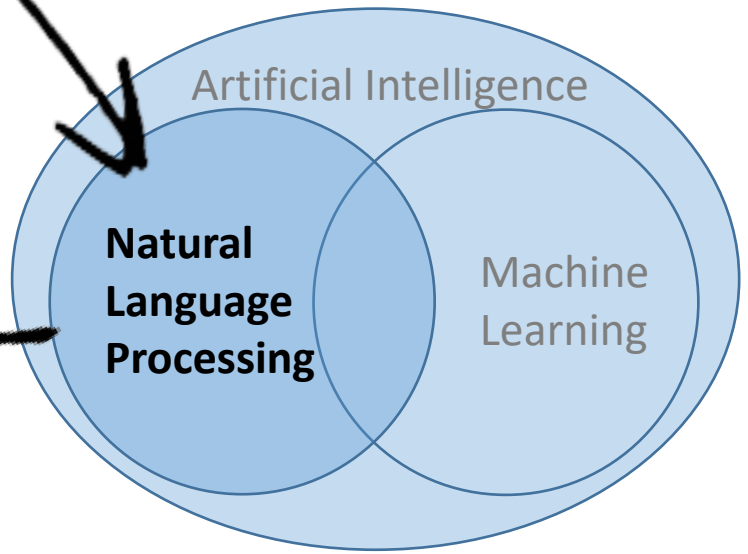
Contextual knowledge

General knowledge

Natural Language Processing (NLP)

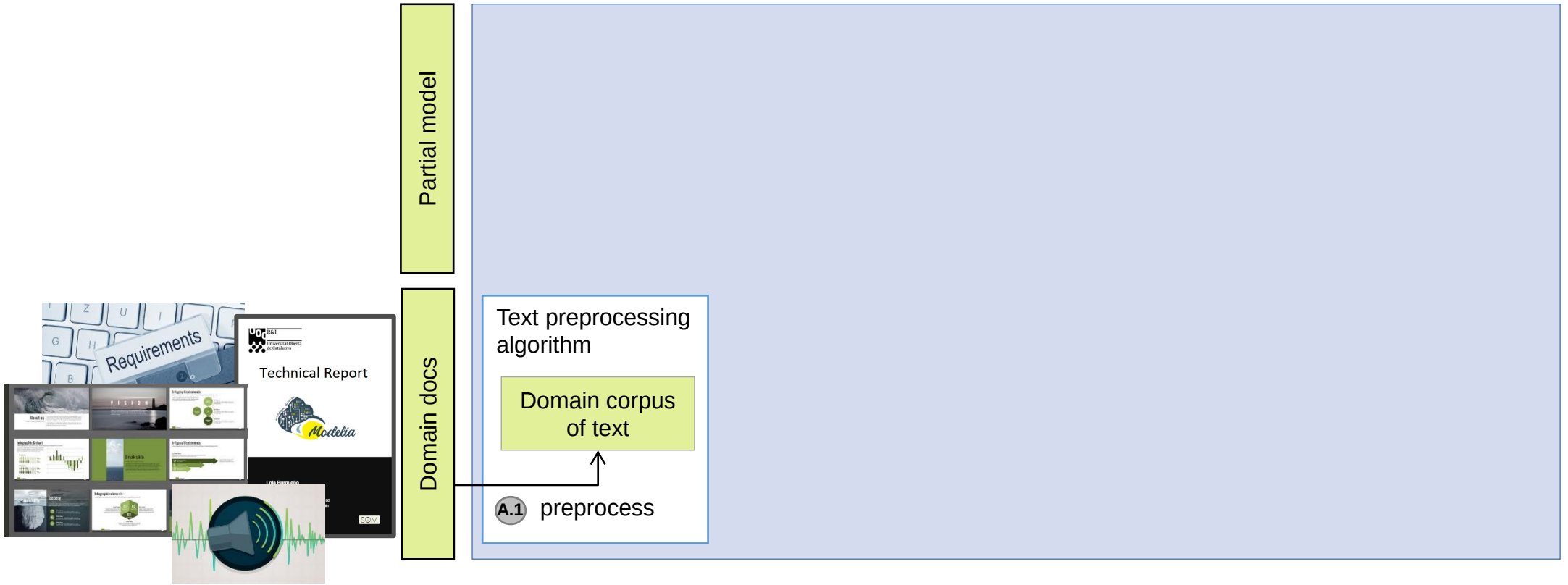






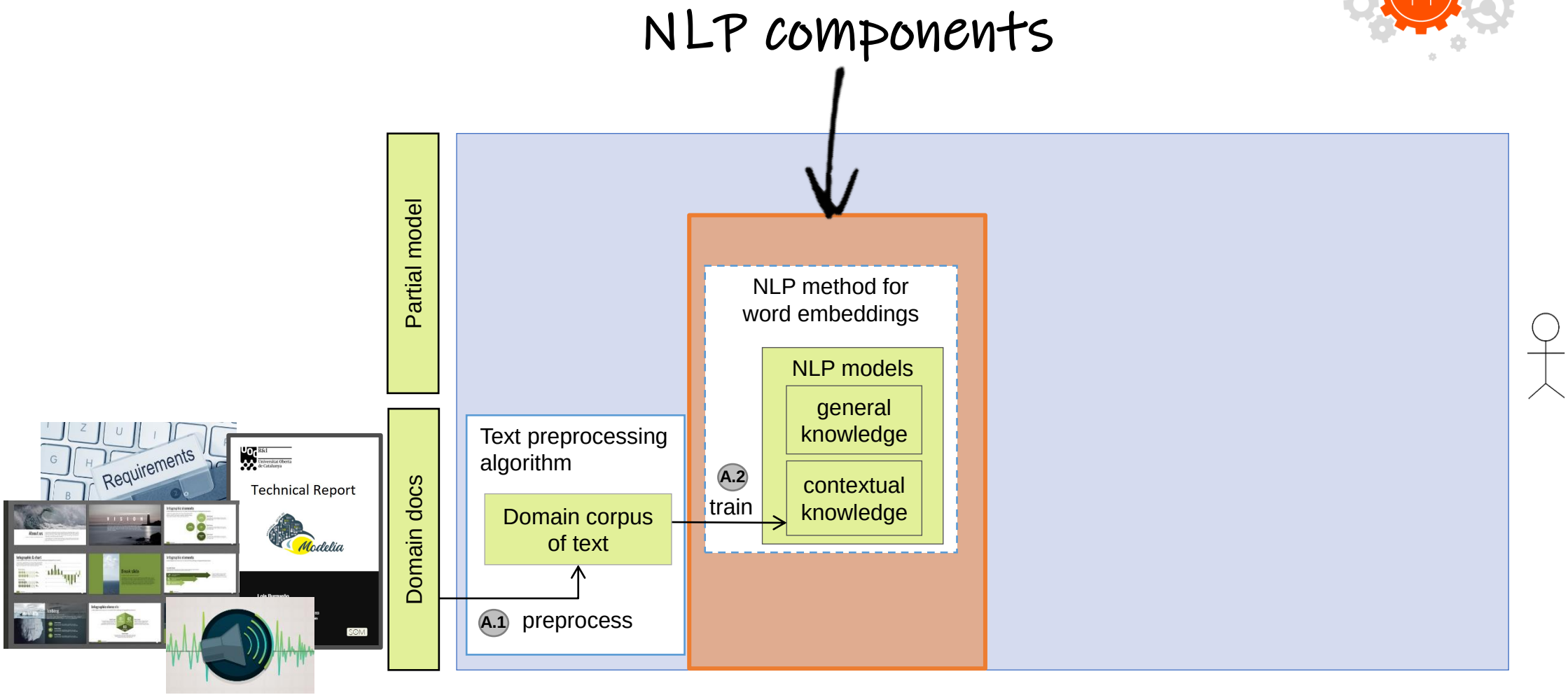
assist

Our approach



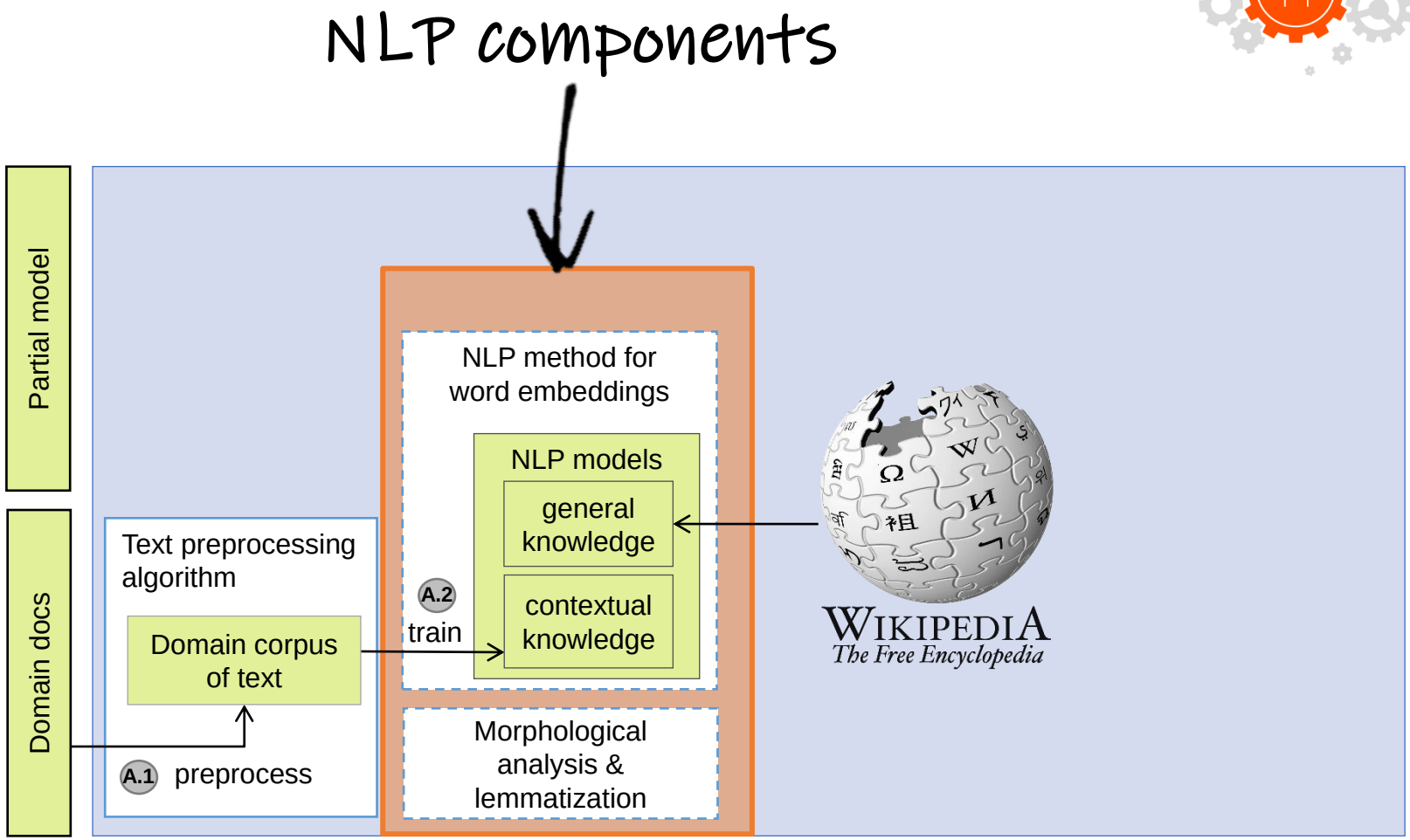
L. Burgueño, R. Clarisó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106

Our approach



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

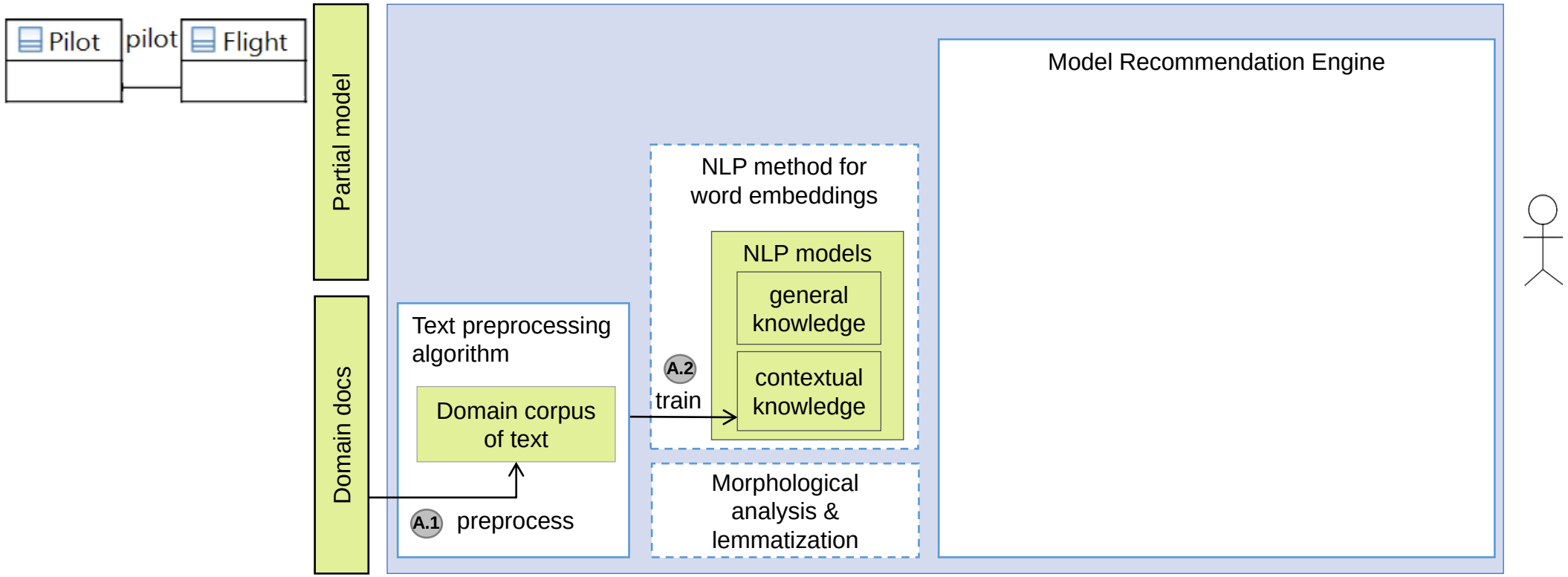
Morphological analysis & lemmatization

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

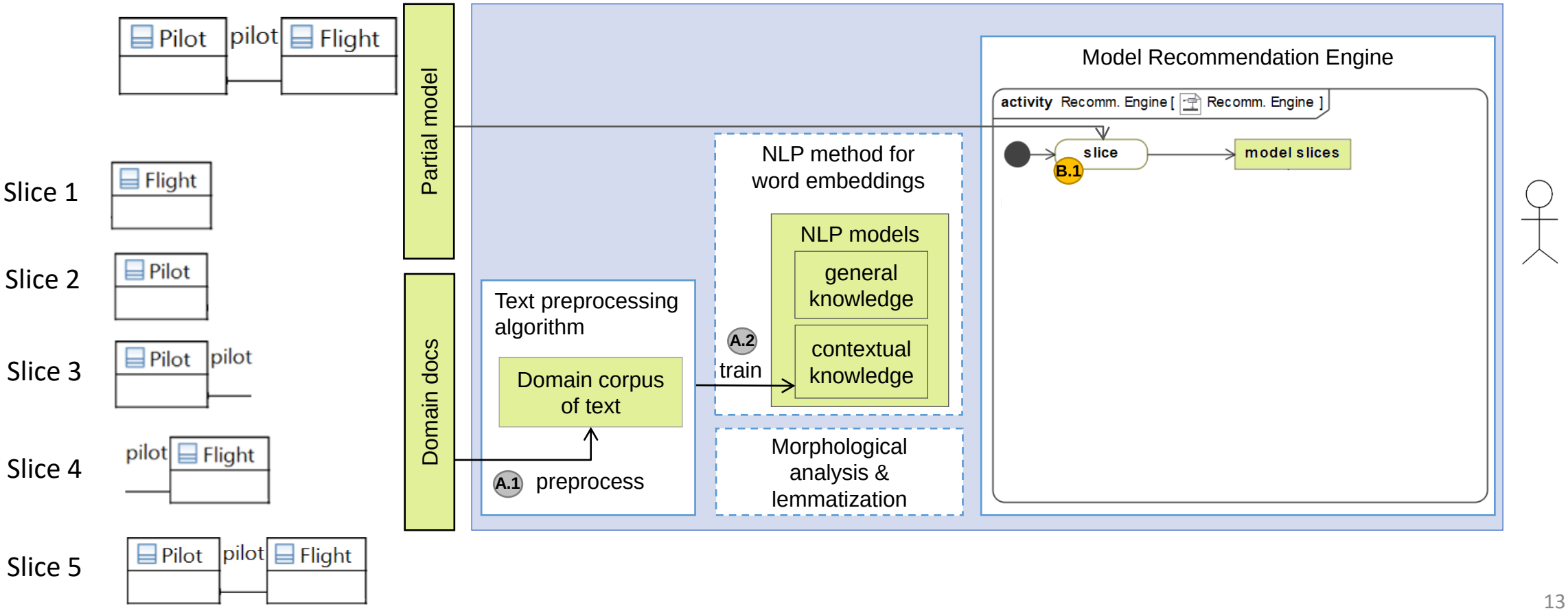
We use WordNet
part of the Python NLTK (Natural Language Toolkit)

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



Our approach

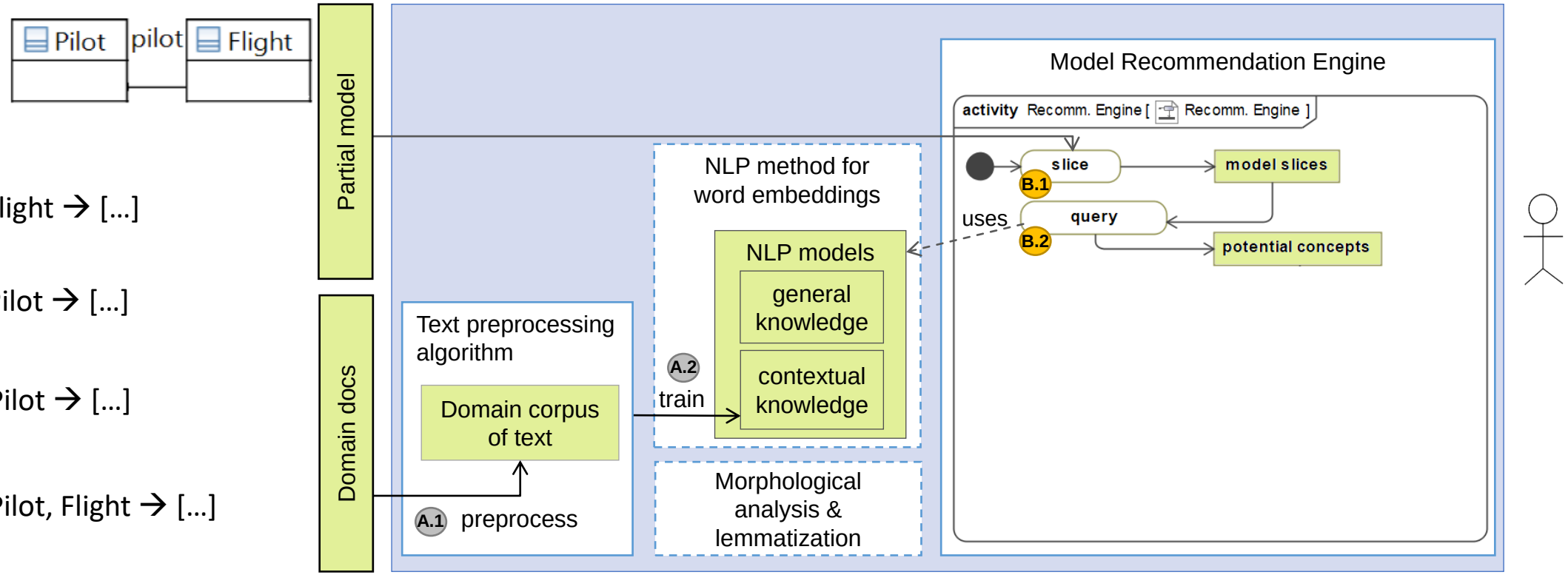
Slice 1: Flight → […]

Slice 2: Pilot → […]

Slice 3: Pilot → […]

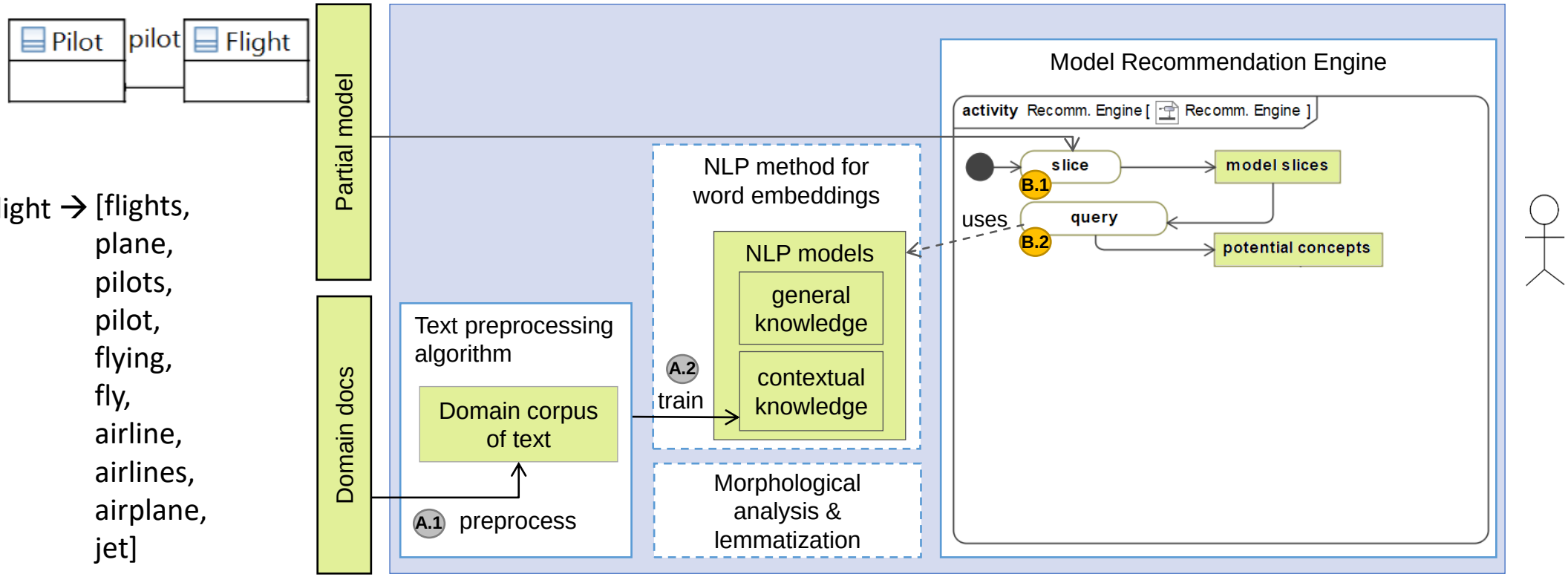
Slice 4: Pilot, Flight → […]

Slice 5: Pilot, Flight → […]



Our approach

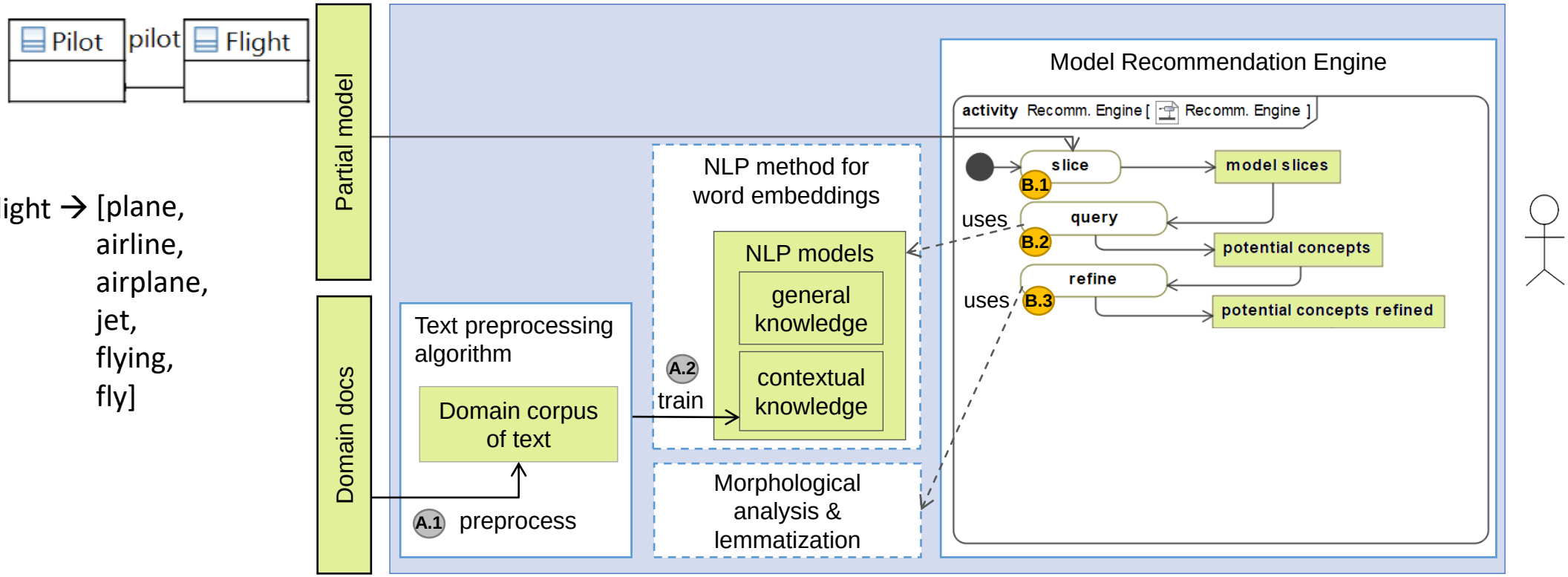
Slice 1: Flight → [flights, plane, pilots, pilot, flying, fly, airline, airlines, airplane, jet]



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

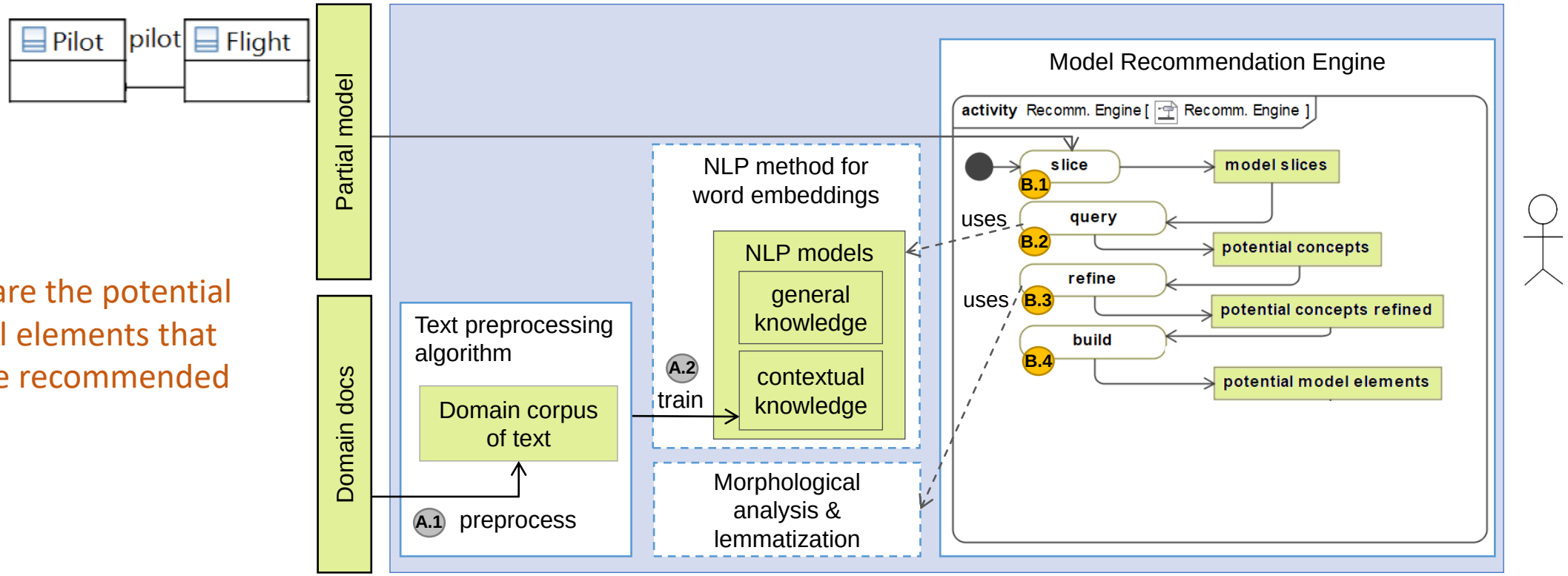
Slice 1: Flight → [plane, airline, airplane, jet, flying, fly]



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

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

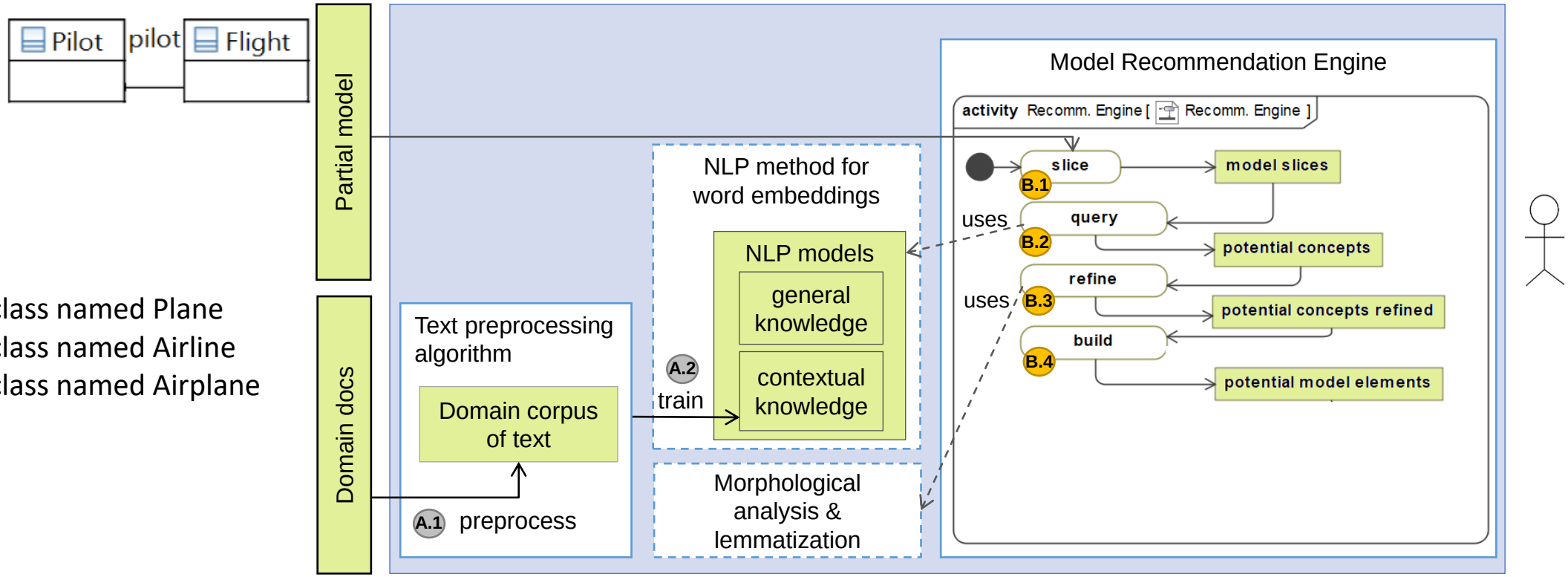
- one slice for each class C in the model (keeping its attributes and dangling relationships)
  → aims to suggest new attributes

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

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

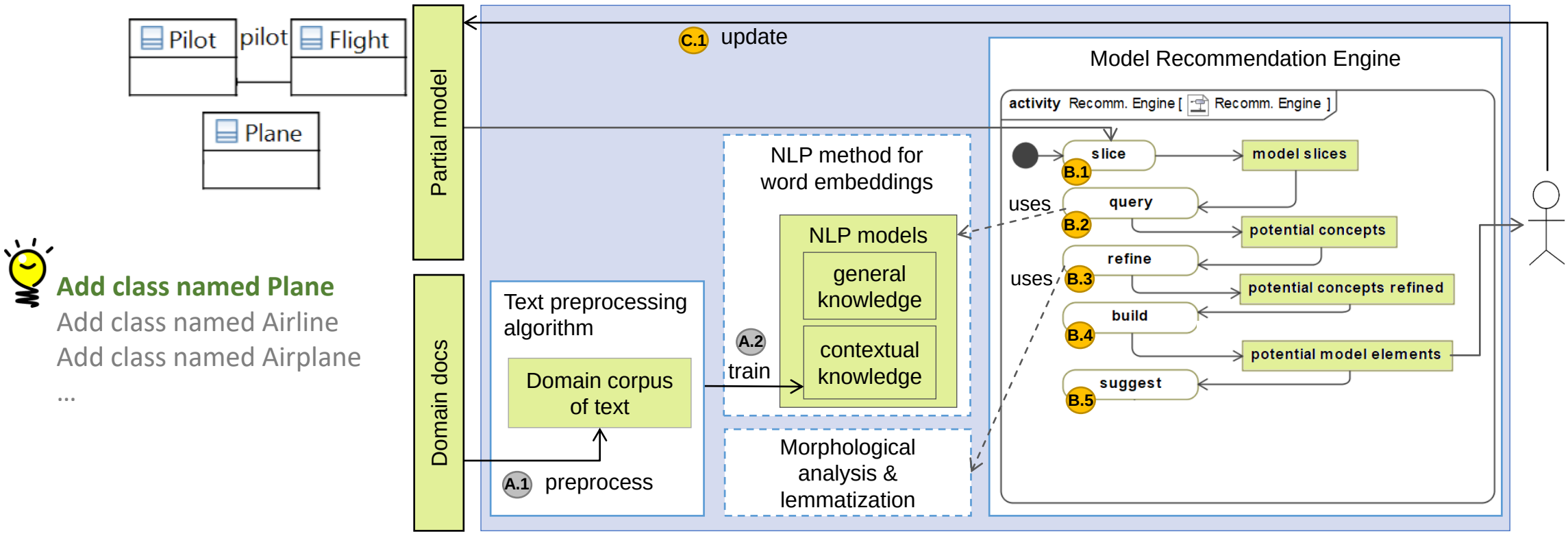
Our approach

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



Add class named Plane

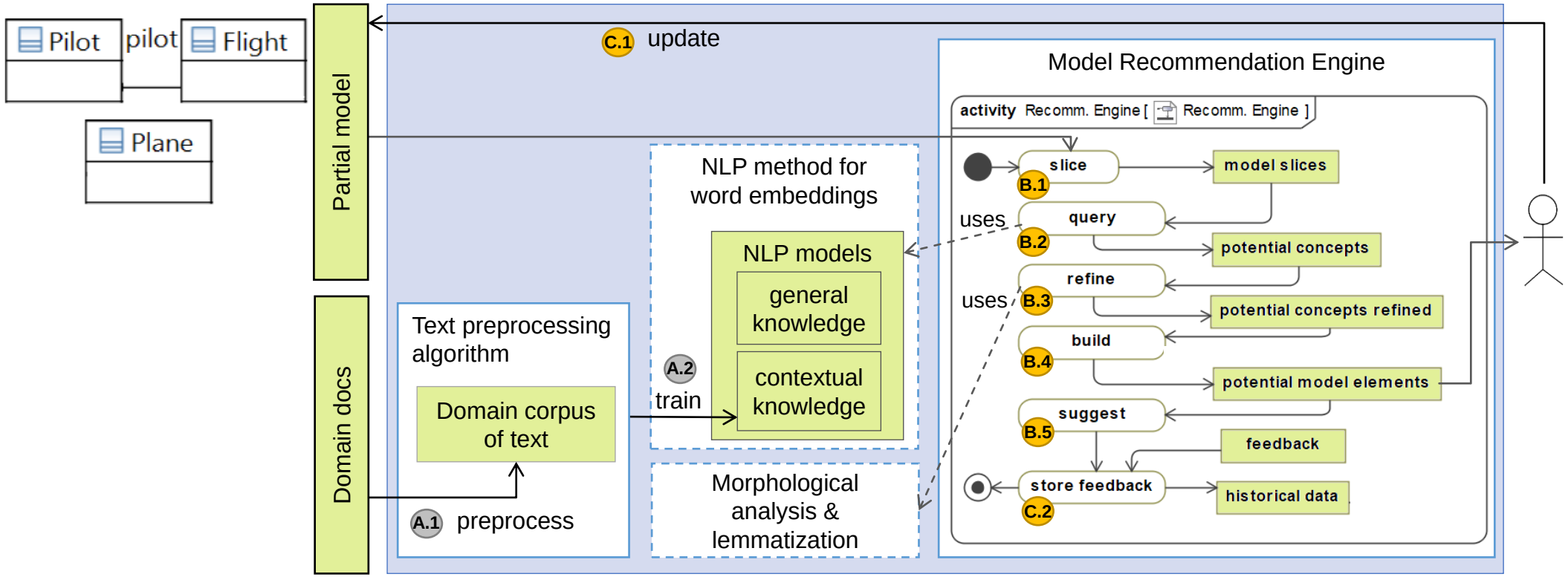
Add class named Airline

Add class named Airplane

...

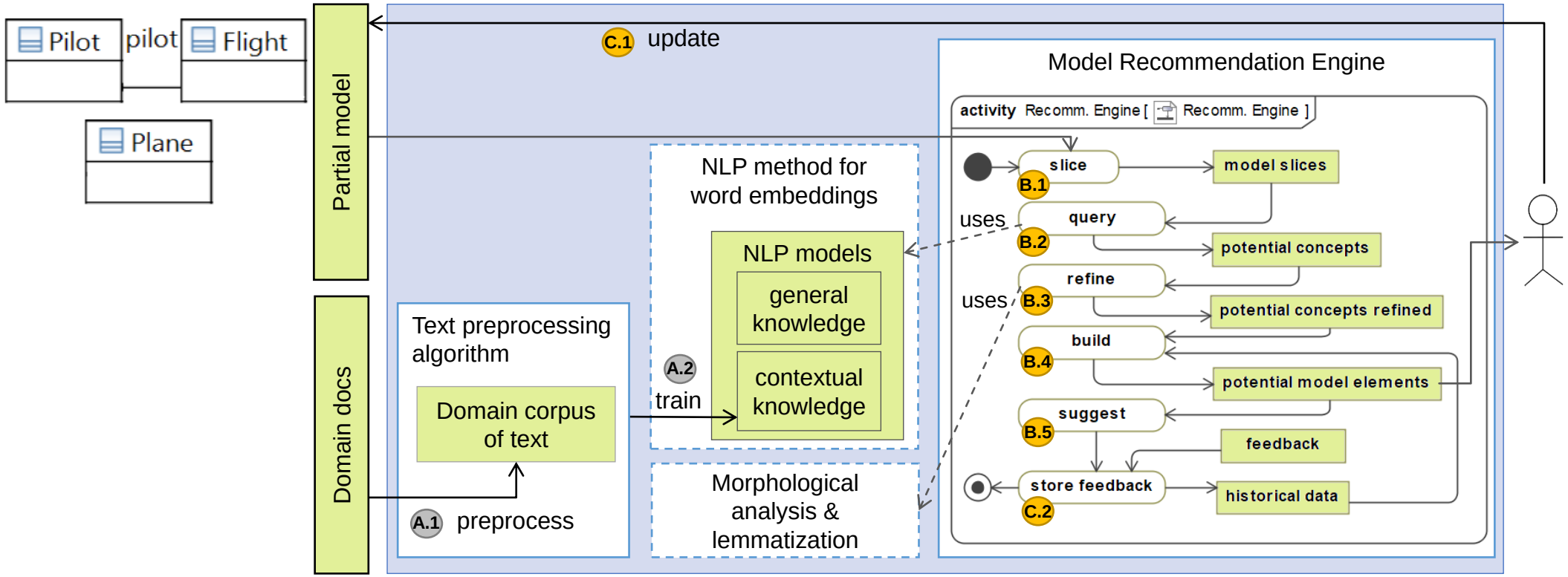
L. Burgueño, R. Clarisó, S. Gérard, S. Li, J. Cabot: An NLP-Based Architecture for the Autocompletion of Partial Domain Models. CAiSE 2021: 91-106

Our approach



L. Burgueño, R. 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. 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

geodes Research Day AND Symposium

Lola Burgueño

Open University of Catalonia, Barcelona, Spain

@LolaBurgueno

lburguenoc@uoc.edu

May 17th, 2022