



PiterPy



TECHNICAL CONFERENCE FOR HARDCORE PYTHON DEVELOPERS

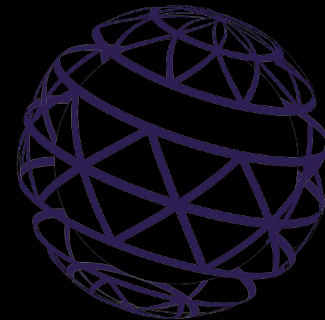


SAINT PETERSBURG
2019 NOVEMBER 1

Learning a Language using NLU/NLG

William Galindez Arias

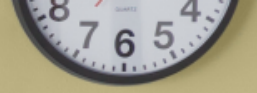
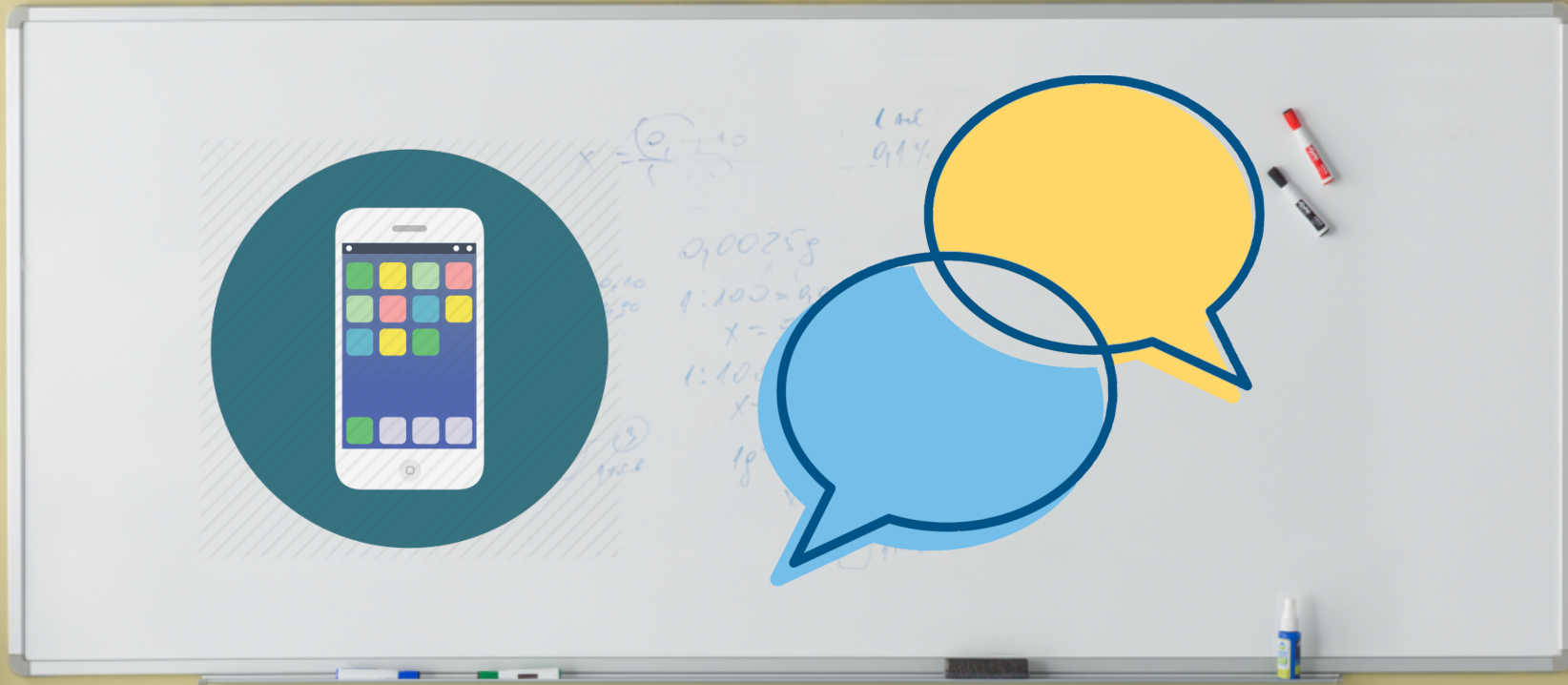
@WilliamAriasZ 



1. Motivation



1. Learn Czech (a bit)





One Direction



**You learn how to say things, but hardly
how to react**



Process the information and reply back

Learn Czech
To survive



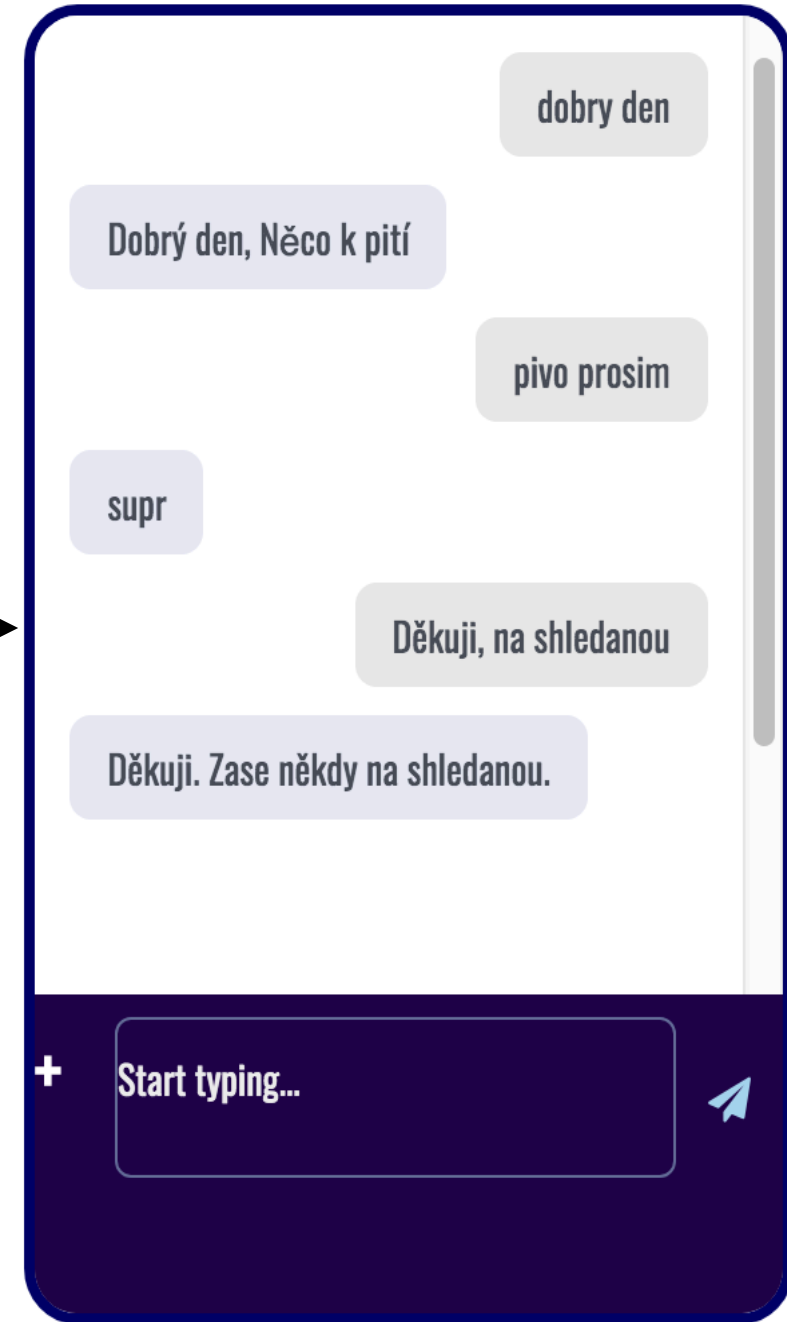
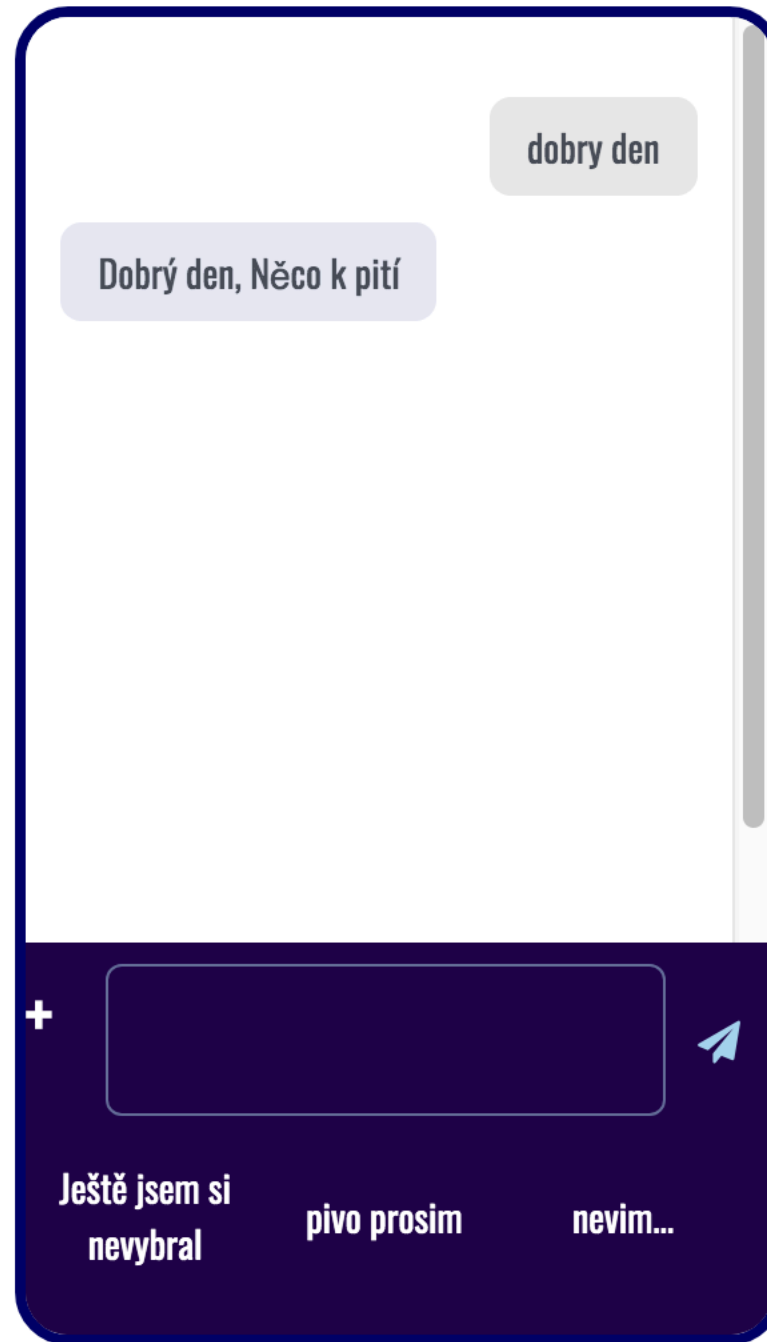
Achieve a Goal

Immediate Impact

What is relevant to your
context ?

Surviving: Asking Beer

Fluent Waiter, many combinations
of utterances (Jabs, uppercuts,
hooks)



Chatbot

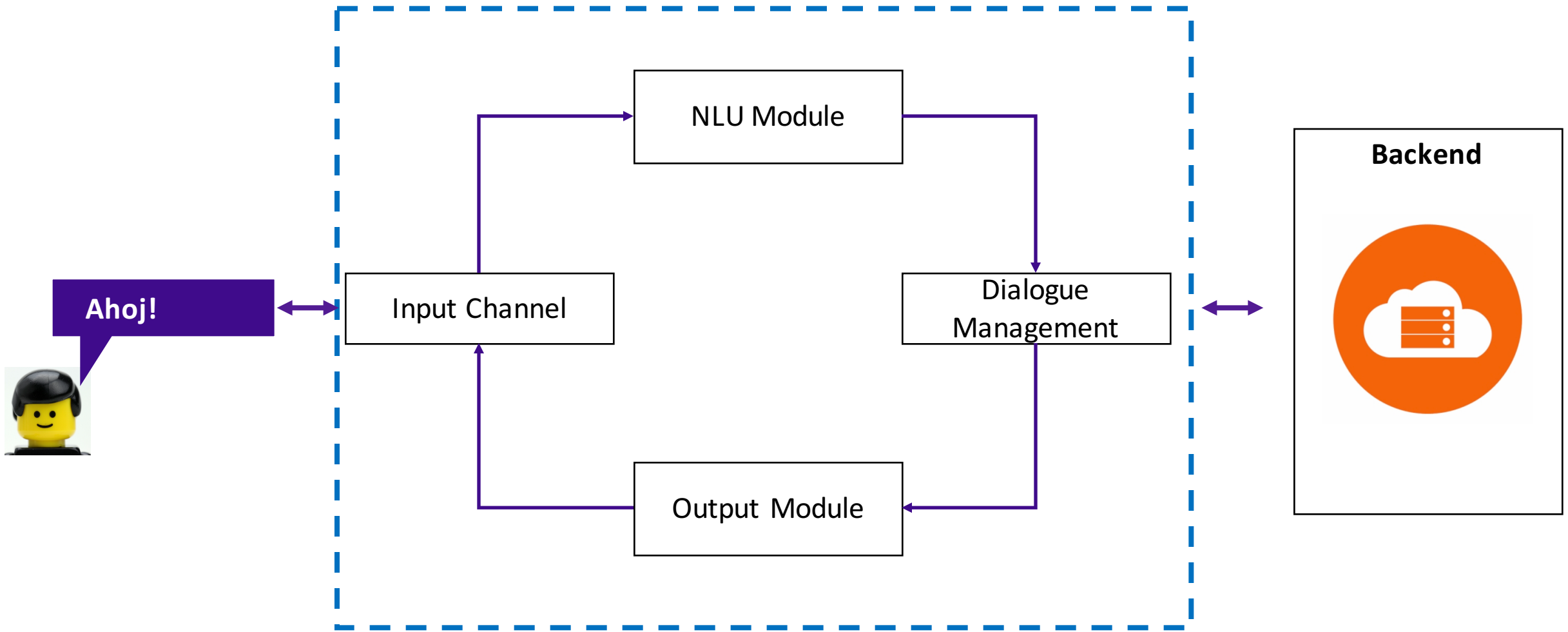
Relevant Context = Closed - Domain
Achieve a goal = Goal-Oriented
Vocabulary = Dataset or Corpus



2. Bot making (New/ Better “Problems”)

~~1. Motivation, Problem statement~~

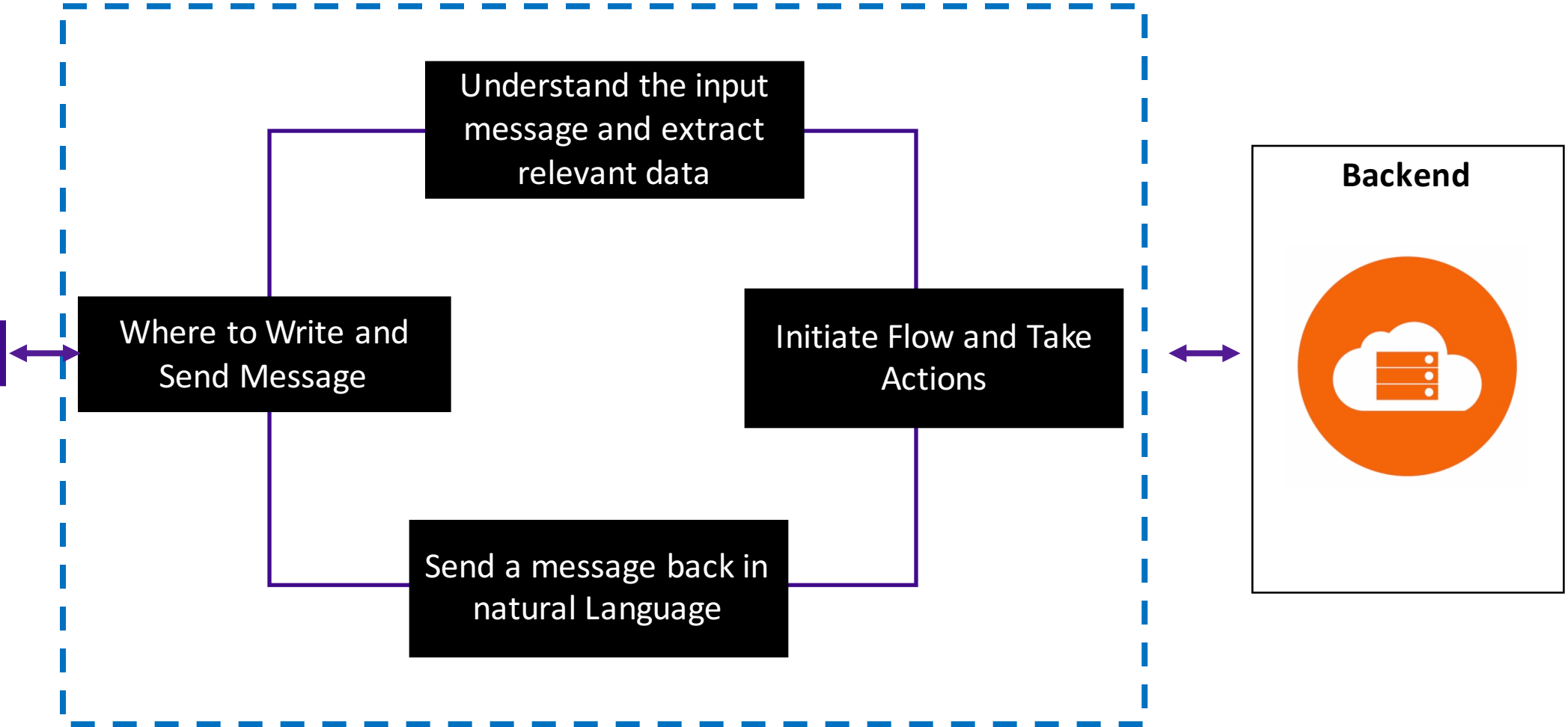
Conversational Software



Conversational Software



Ahoj!



Bot Development Framework

Rasa

How can I train NLU model without initial word vectors?



- William (Someday in December 2018)

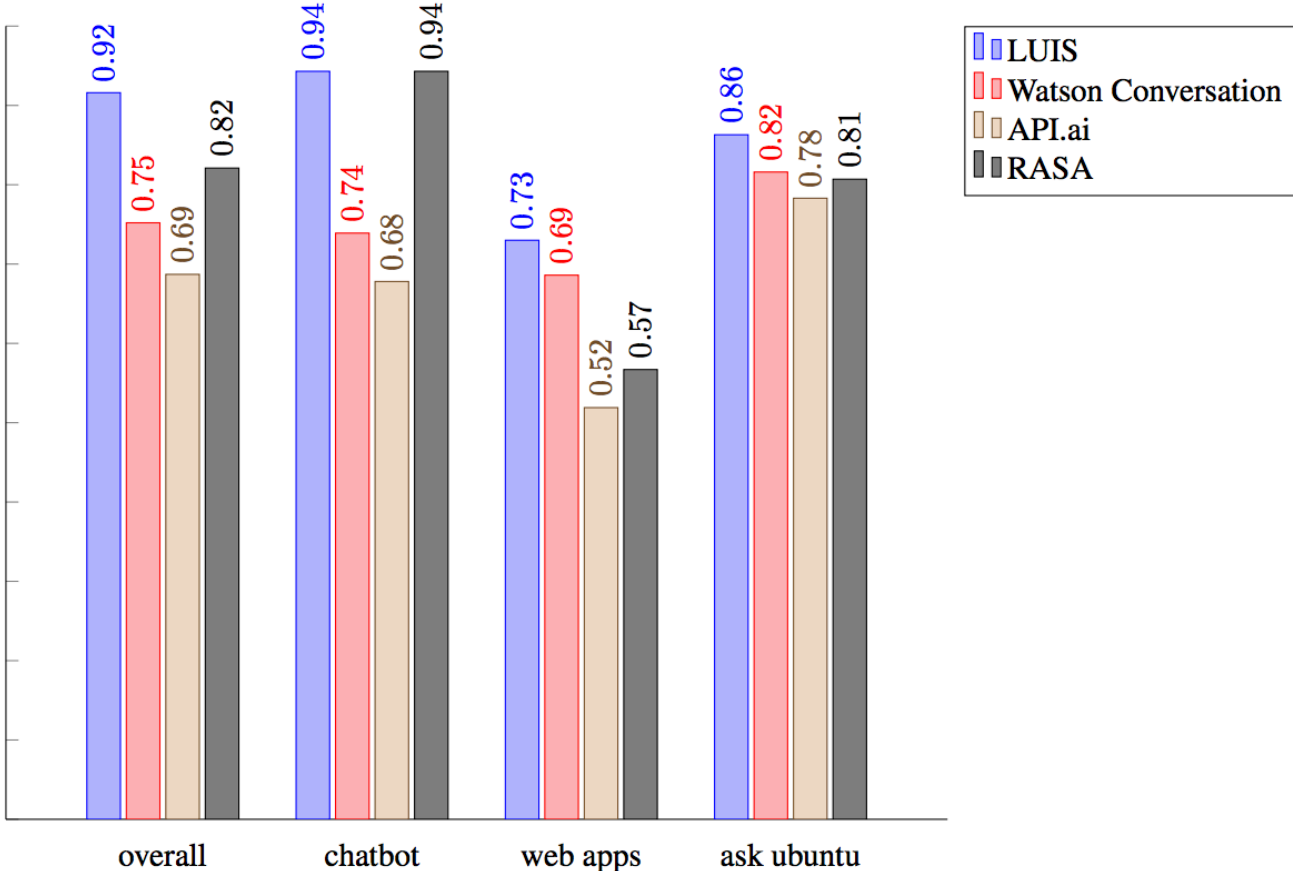


Figure 3: F-scores for the different NLU services, grouped by corpus

Evaluating Natural Language Understanding Services for Conversational Question Answering Systems

<https://www.aclweb.org/anthology/W17-5522.pdf>

Intent: Eat

Entity: Ham



Natural Language Understanding

Czech



Dám si jedno
pivo Pilsener

NLU

Intents

ask_table

order_drink

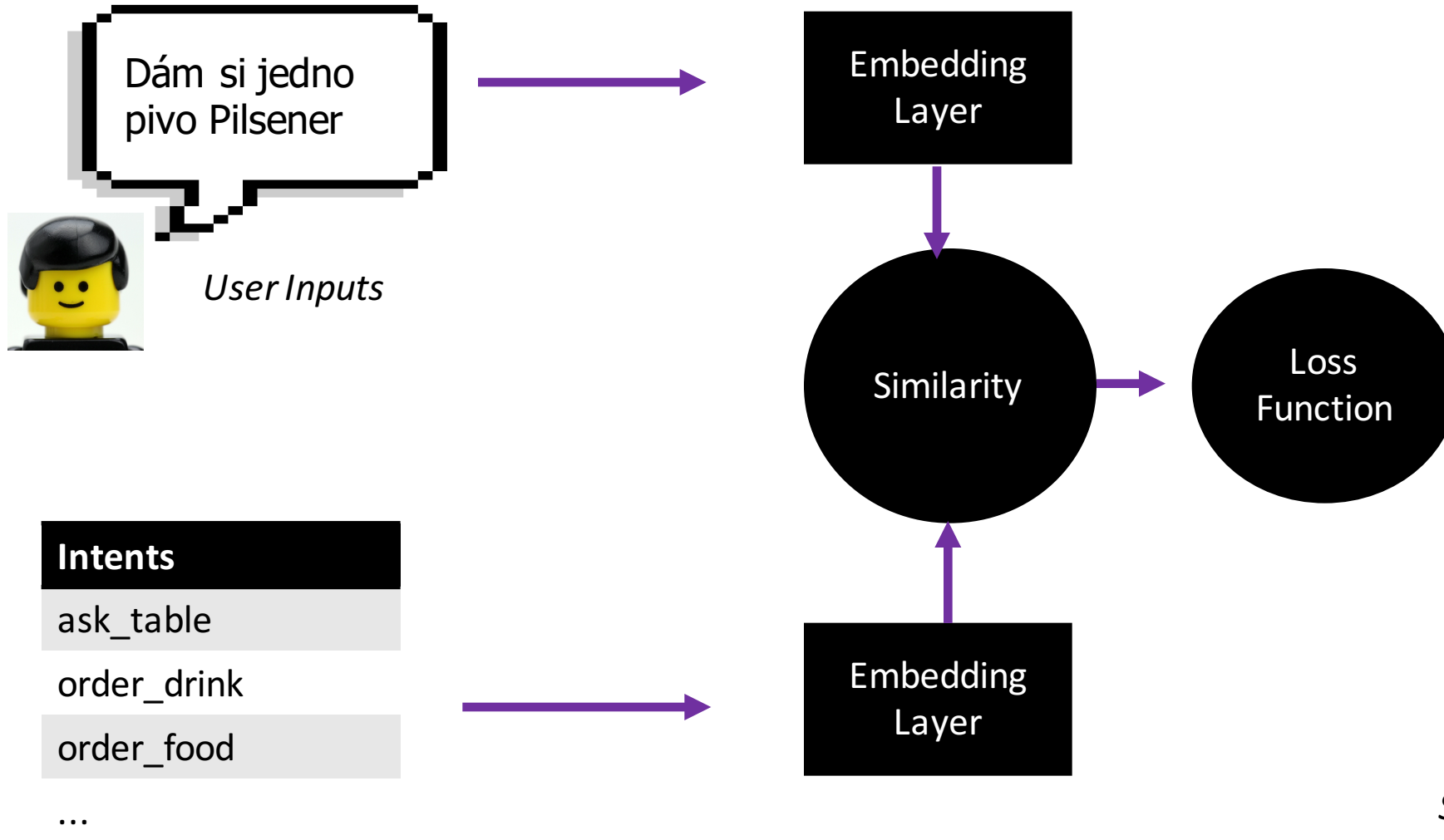
order_food

...

```
{'drink': pivo,  
'type': pilsner}
```

NLU: Word Vectors from Scratch

Czech

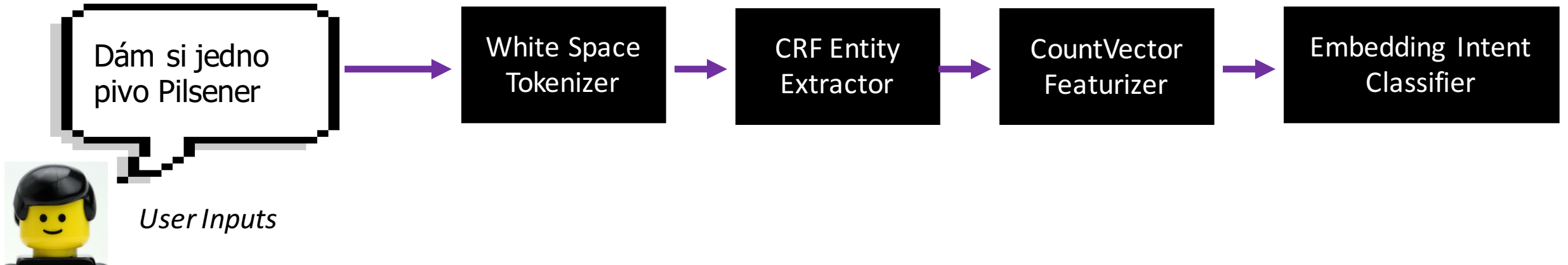


Starspace: Embed All the things!

<https://arxiv.org/abs/1709.03856>

NLU: Word Vectors from Scratch

Pipeline used for Czech



```
print(interpreter.parse("Chtěl bych si objednat Smažený sýr"))
```

```
{ 'intent' :  
{  
'name' : 'order_food_main_dish',  
'confidence' : 0.8891581892967224},  
'entities' : [{ 'start' : 23, 'end' : 34,  
'value' : 'Smažený sýr', 'entity' : 'dish',  
'confidence' : 0.6718105031002314,  
'extractor' : 'ner_crf'  
}]  
}
```

No pre-trained word Vectors in your language?

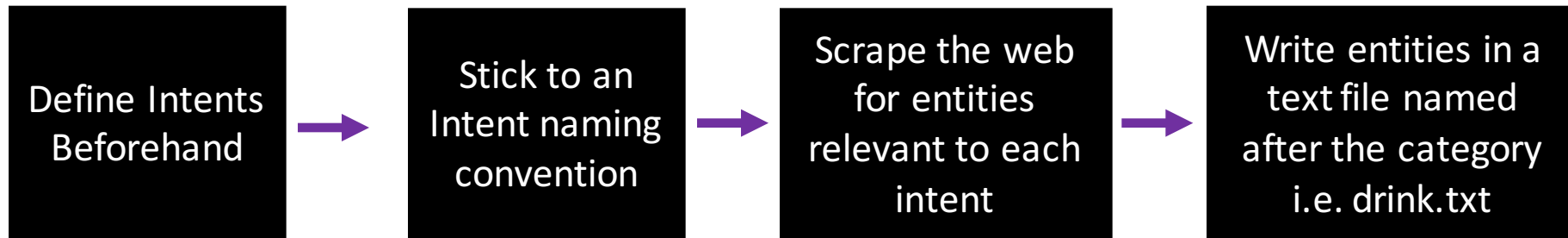


Very Specialized vocabulary in your domain?



Obtaining Data to train NLU Model

Czech

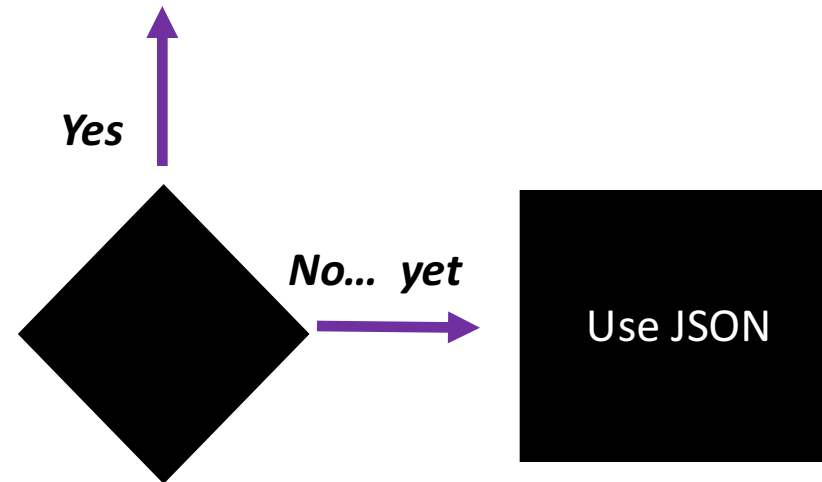


How I did it



Training the Model – Data Format

Czech



Do you want non-technical users feed your training data?



Training the Model – Data Format

Czech

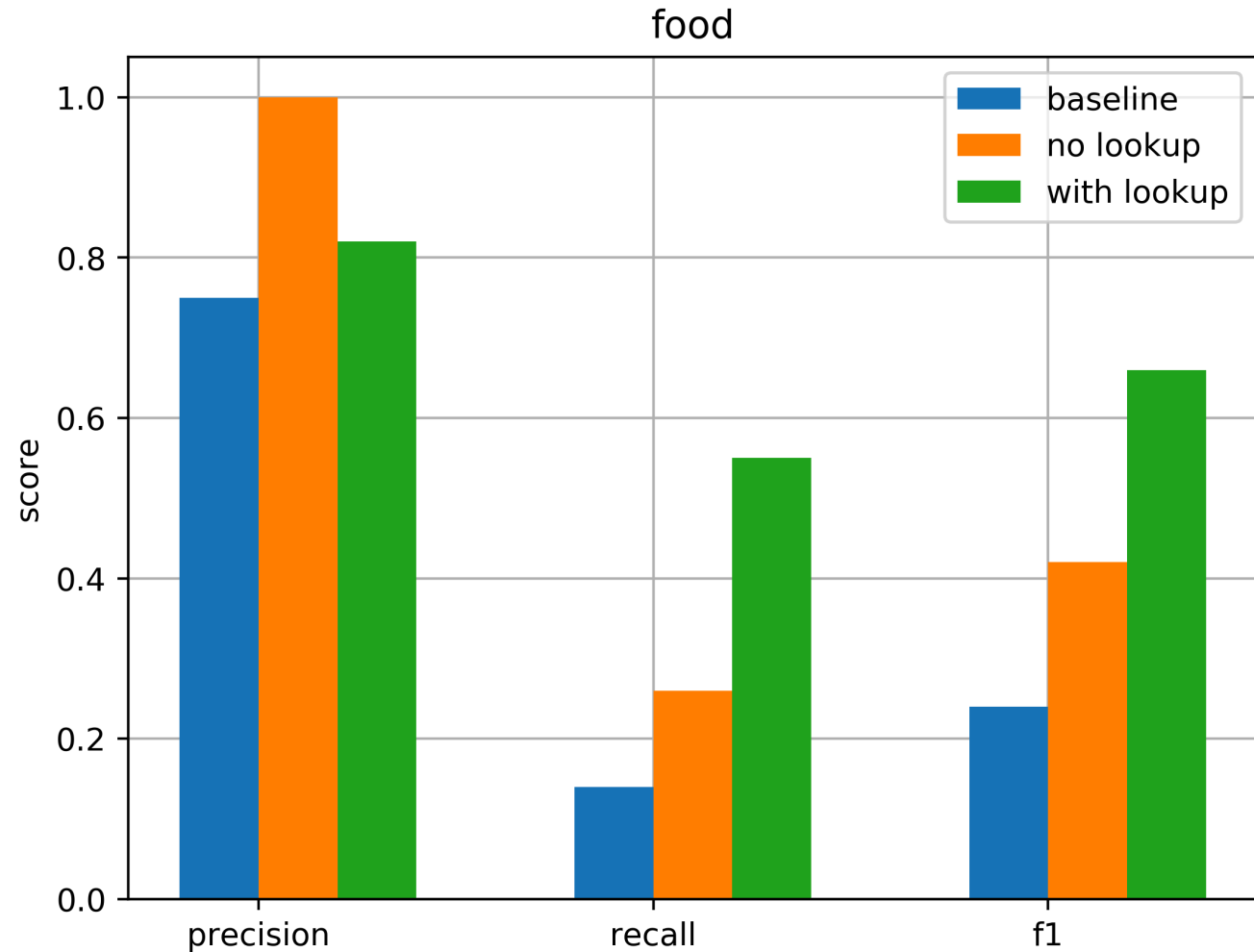
```
## intent:order_drink
- Mohl bych dostat sklenku [perlivé vody](drink)?
- Mám rád [pivo](drink).
- Chutná mi [minerálku](drink).
- Chtěl bych sklenku [vína](drink)
- chci [pivo](drink)
- dal bych si [pivo](drink)
- Chtěl bych [vodu](drink)
- Chci [vino](drink)
- Dám si jedno [pivo](drink) prosím
```

This is how the training data might look



Training the Model – Using Look Up Tables

Czech



Clean and filter your
Lookup table in advance

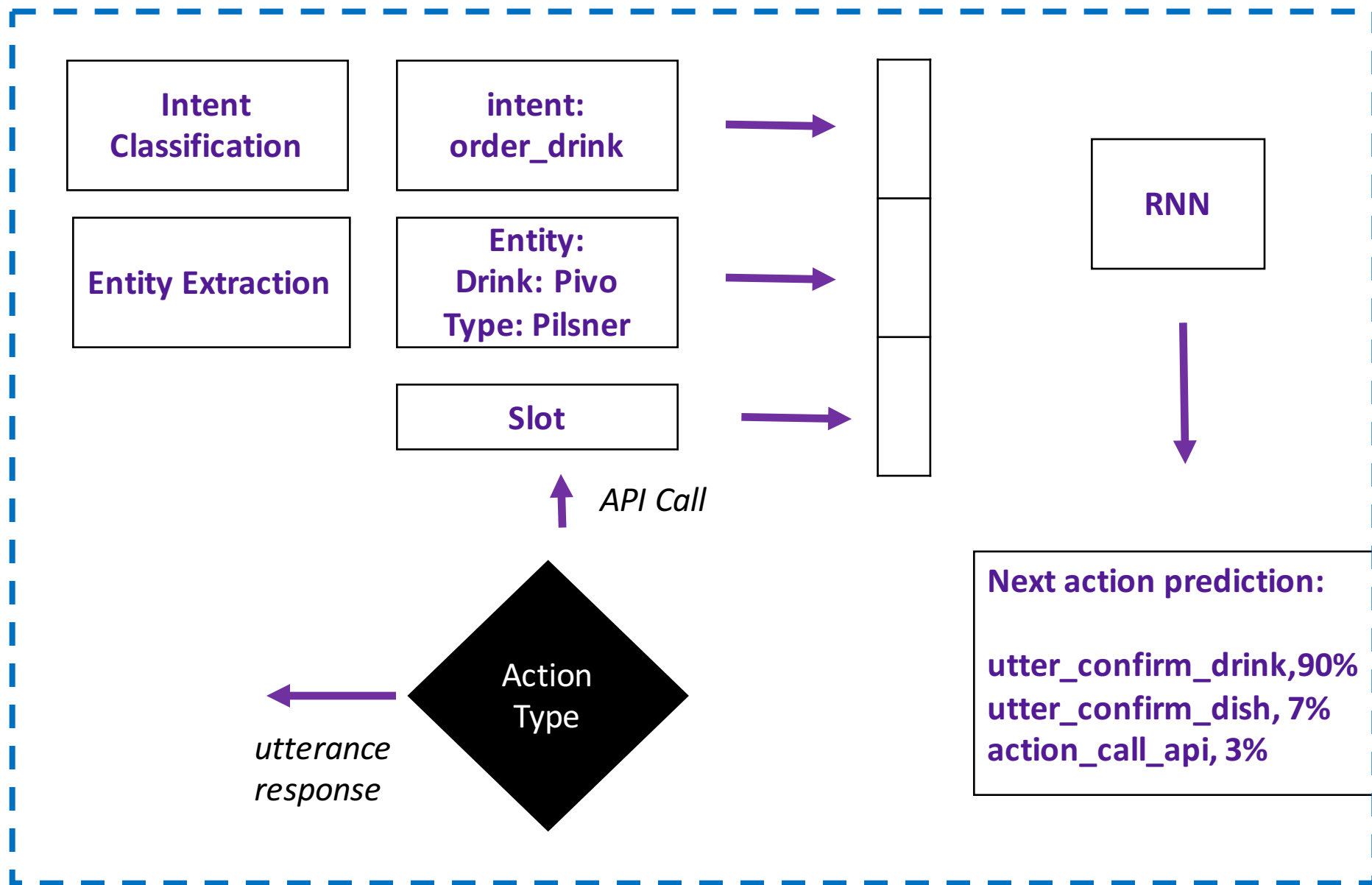


<https://rasa.com/docs/rasa/nlu/training-data-format/#lookup-tables>

[https://medium.com/rasa-blog/entity-extraction-with-the-new-lookup-table-feature-in-rasa-nlu-](https://medium.com/rasa-blog/entity-extraction-with-the-new-lookup-table-feature-in-rasa-nlu-94c6c30876a3)

[94c6c30876a3](https://medium.com/rasa-blog/entity-extraction-with-the-new-lookup-table-feature-in-rasa-nlu-94c6c30876a3)

Dialogue Czech



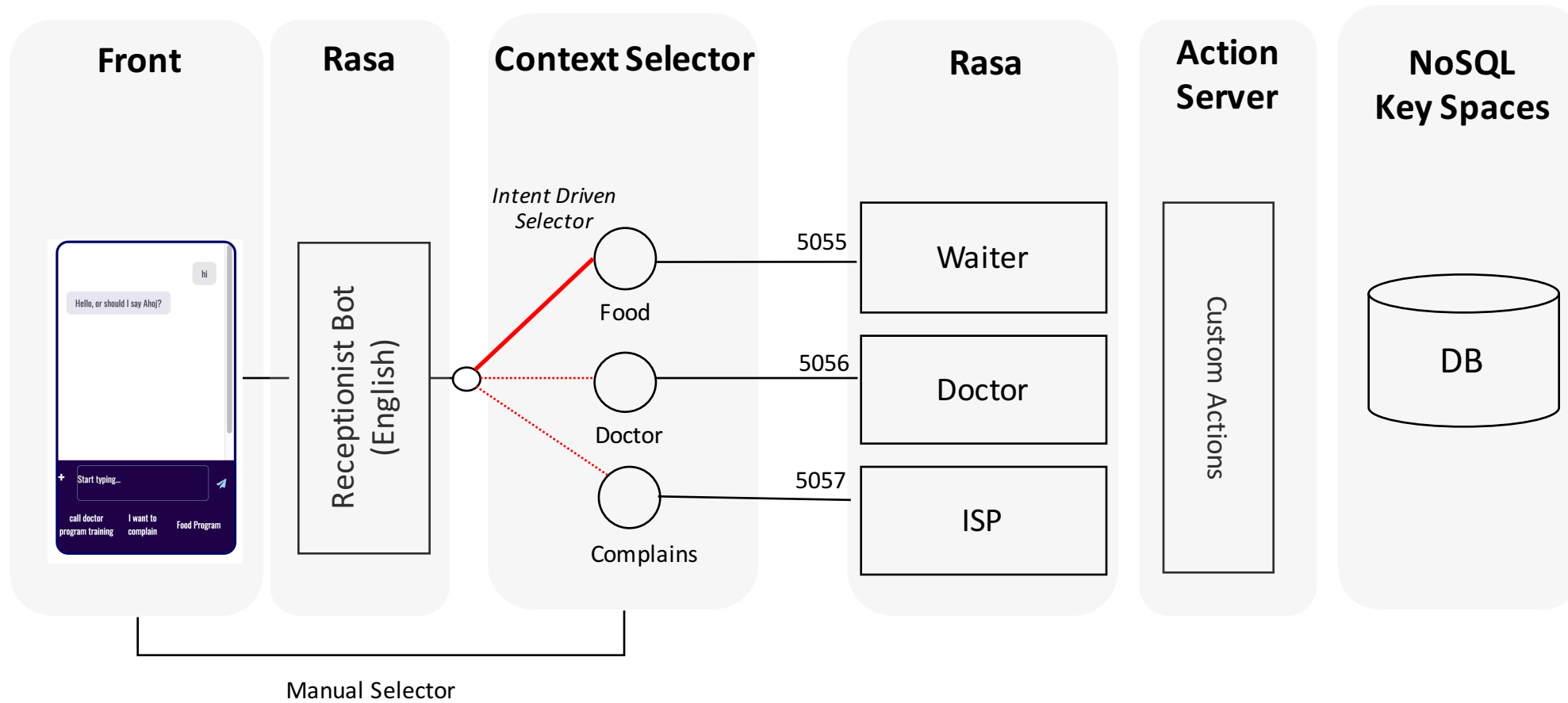


3 Language beyond my use case

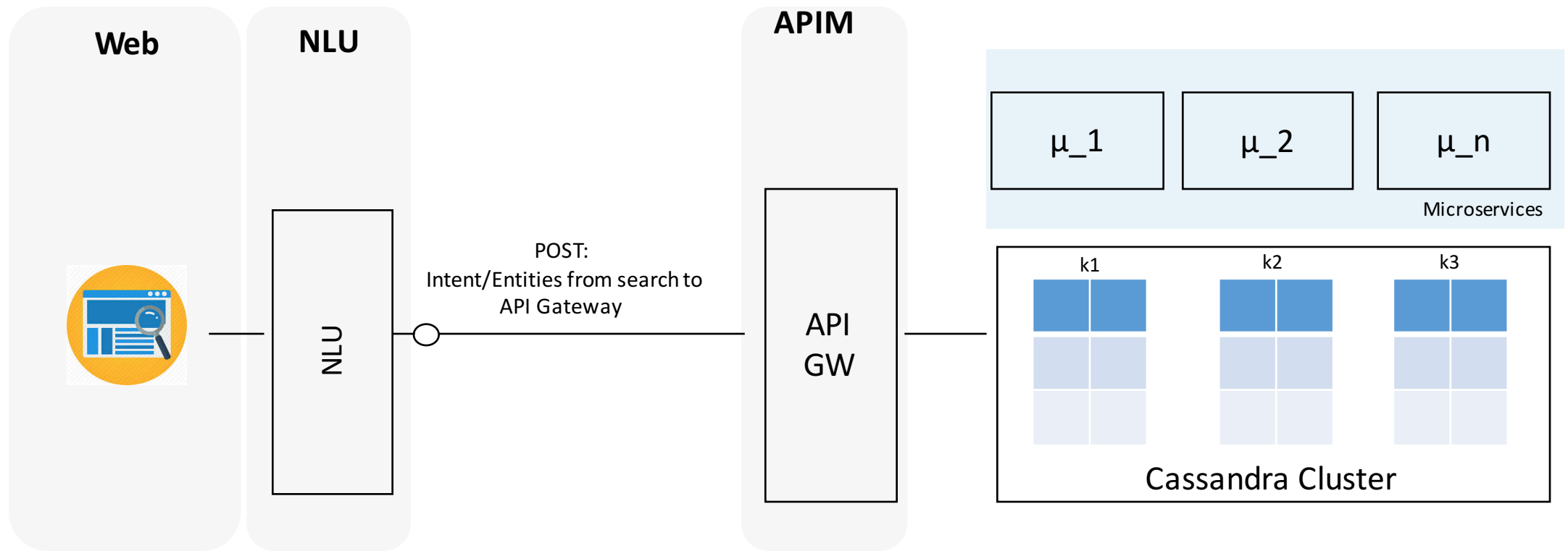
~~2 The making (New/ Better Problems)~~

~~1. Motivation~~

Architecture v1.0



Architecture for Searching using NL



Problem: Time consuming to find Staff in legacy tools

Proposed Solution: Search with natural queries:

"Find me candidates with Python Skills available to start next week"

No word Embeddings? Learn your own and relevant to your domain.

(Start small, and should work in any language)

Retrain Based on real conversations

Experiment with BERT

</end>