inovex

Berlin Buzzwords

Querying Elasticsearch with Deep Learning to Answer Natural Language Questions

Sebastian Blank, Hans-Peter Zorn

Berlin, 2019-06-17

Sebastian Blank

Data Scientist @ inovex

@SbstnBlank

Hans-Peter Zorn
@data_hpz

Dr. Florian Wilhelm
@FlorianWilhelm

Prof. Dr. Achim Rettinger
Universität Trier

Agenda

1. Use Case
2. Approach
3. Results
4. Learnings

Use Case

Voice Control will shape our lifes

“You will be able to do pretty much anything via voice command.”

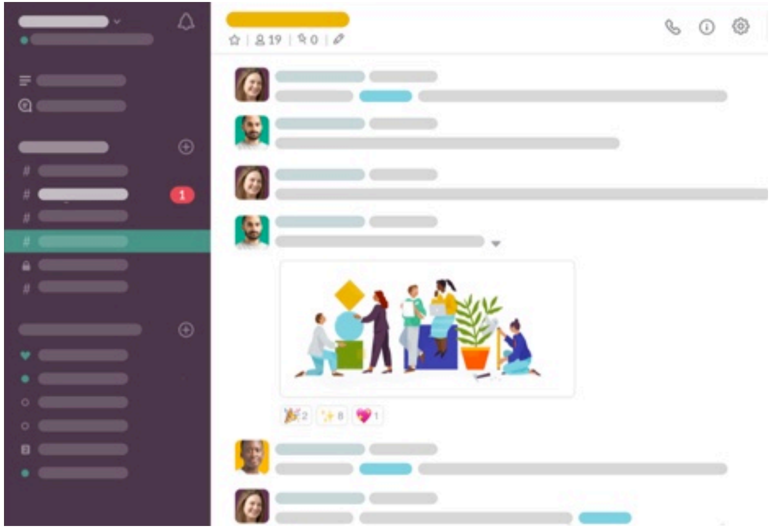
Elon Musk about Teslas Model 3

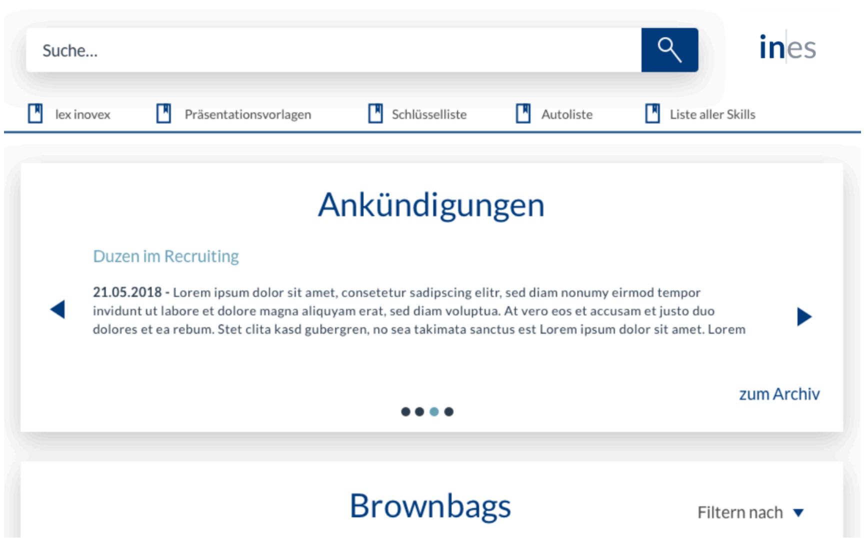
„Speech is going to be the interface at home.“

Kenn Harper,
Nuance Communications

Images from Google Image Search

How it started





Images from Google Image Search

Conversations require background information

Who is the president
of the United States?

Which customers received a
coupon and used it in our shop?

Who starred in Avatar?

What appointments
do I have tomorrow?

Image from Google Image Search

Query languages impede access to information

Leveraging DL to overcome this barrier

› Hard lookup

› Soft lookup

Hard lookup

inovex

How to access a database with natural language?

- Hard lookup
  - interpretable (+)
  - existing API (+)

  - **non-differentiable (-)**
  - not end-to-end trainable (–)
  - labelling is costly (-)

inovex

Soft lookup

Soft lookup

inovex

How to access a database with natural language?

- Hard lookup
  - interpretable (+)
  - existing API (+)

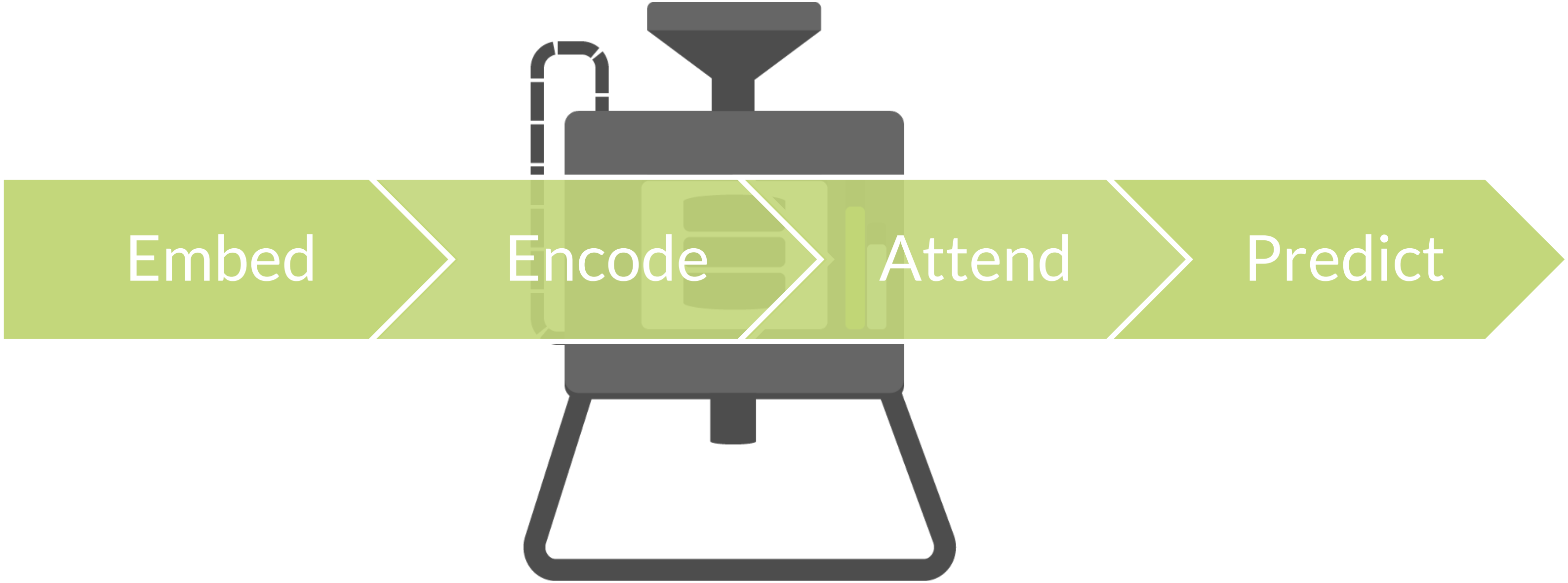
  - non-differentiable (-)
  - not end-to-end trainable (–)
  - labelling is costly (-)

- Soft lookup
  - end-to-end trainable (+)

  - hard to interpret (-)
  - impeded by security & privacy issues (-)
  - capacity (-)

Approach

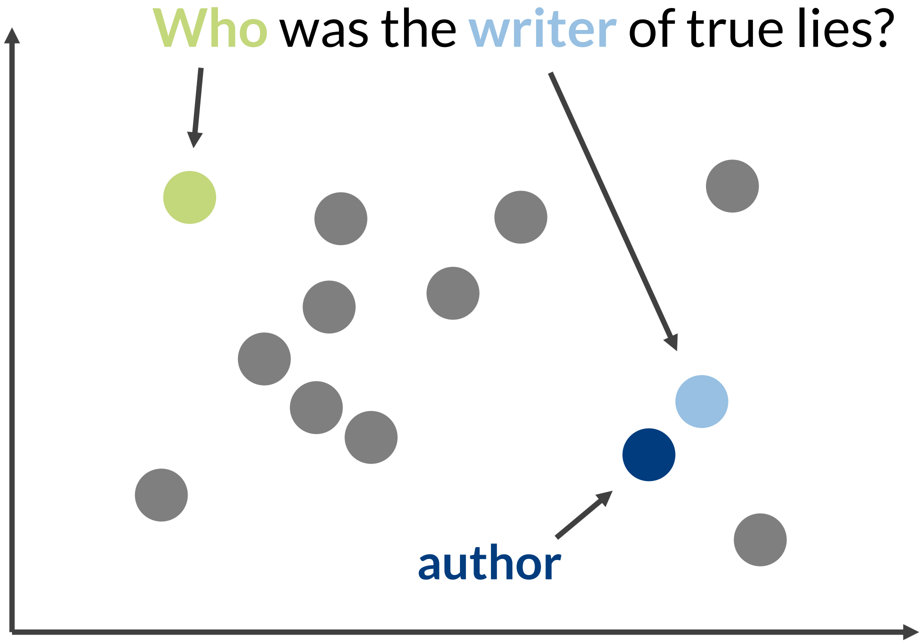
Let‘s build this machine



inovex

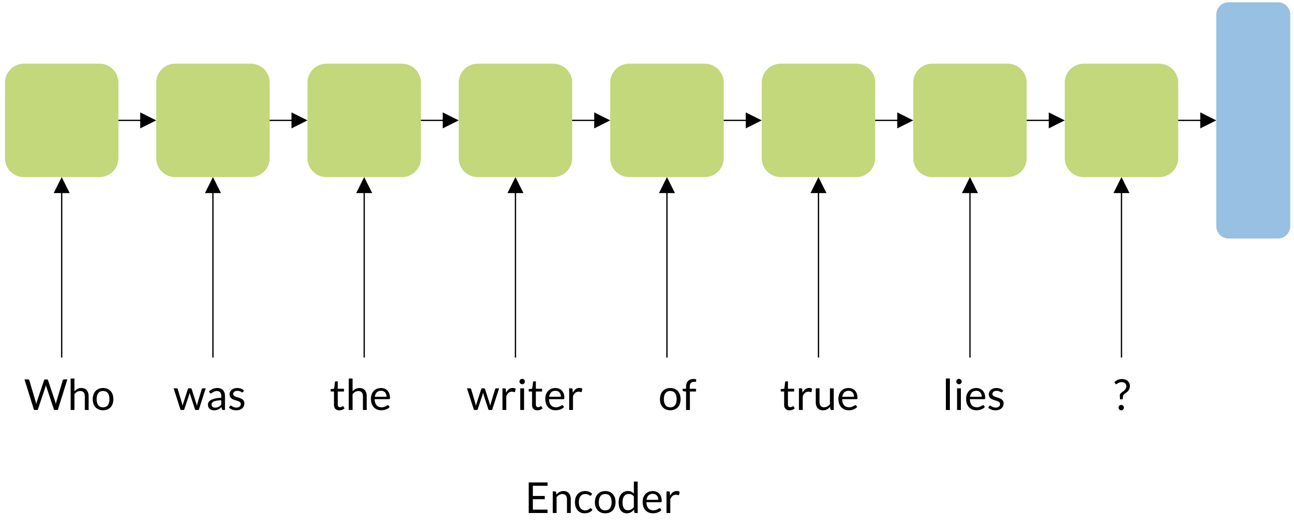
Embedding

Representing words as continuous vectors



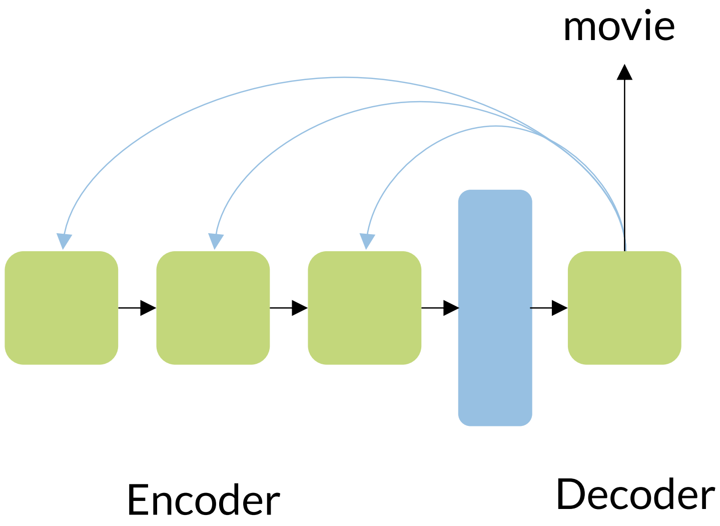
Encoder

Creating a context representation of a sequence



Decoding with Attention

Focussing on important subsets of the input



Pointer Network

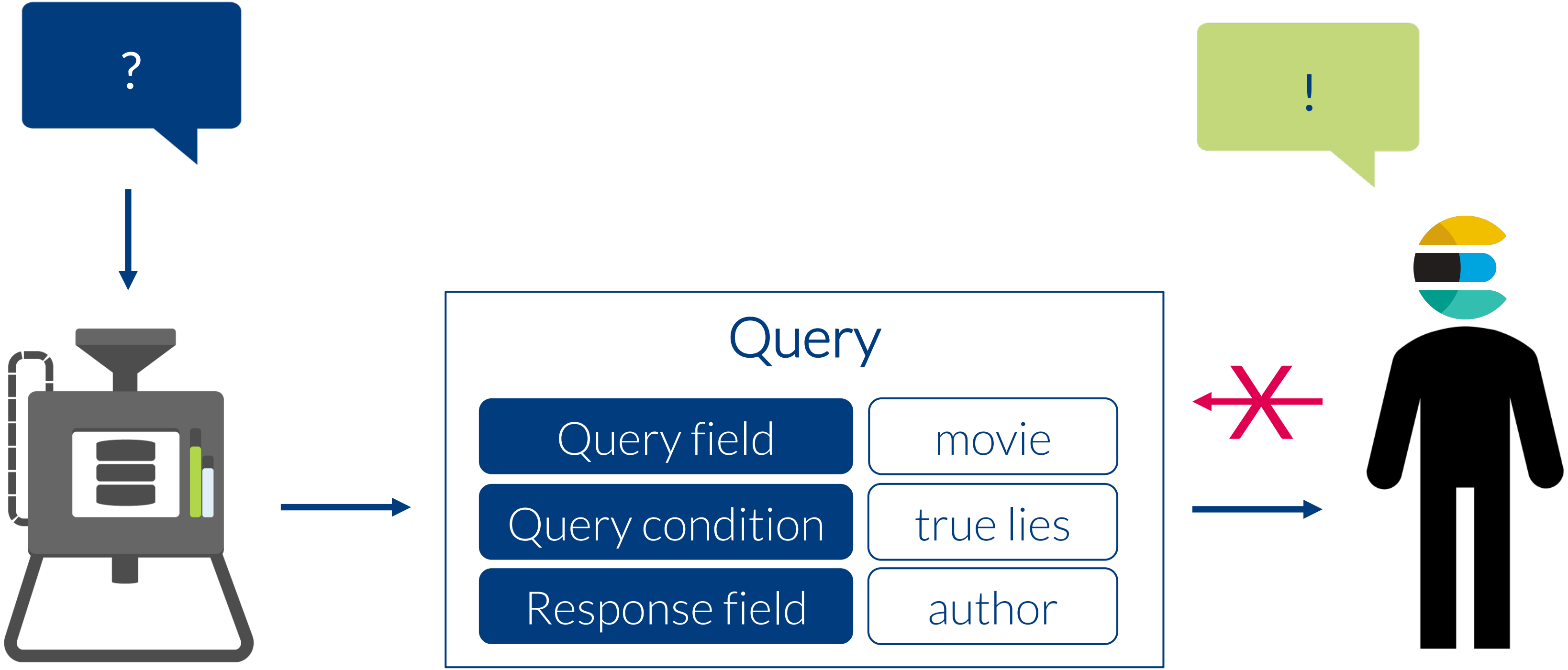
[…] where v, W_s and W_t are trainable parameters and a decoder hidden state h_t is scored against an encoder hidden state $\bar{h}_s$. Pointer attention significantly decreases the output space and therefore reduces […]

Training Procedure

For our benchmark, we implement PointerNet with a bidirectional two-layer LSTM as encoder and a uni-directional two-layer LSTM as decoder, where all recurrent layers consist of 100 units. […]

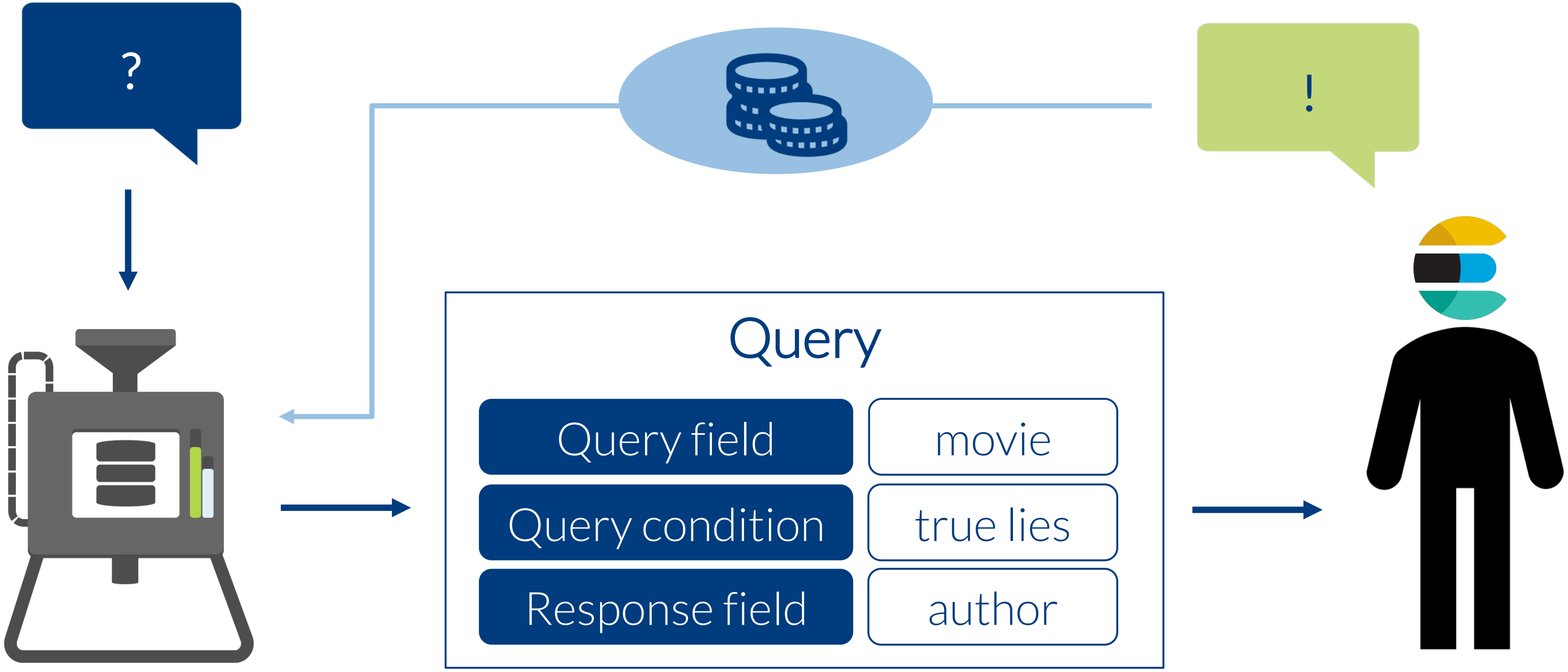
Filling the query template

Non-differentiability impedes end-to-end training



Filling the query template

Reinforcement Learning for the rescue



Rewards

-2

invalid queries

-1

valid queries

incorrect results

+1*

valid queries

correct results

* varied in experiments

Results

Design of the reward function matters

Improved performance due to higher positive rewards

valid queries

correct results

99.0 %

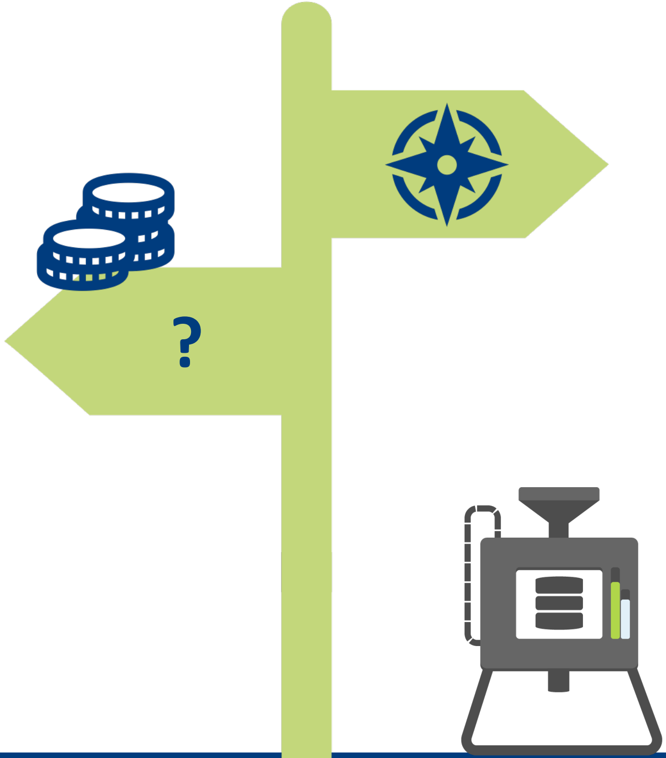
0.0 %

83.3 %

54.5 %

Design of the reward function matters

Exploration boni improve performance



without reward bonus

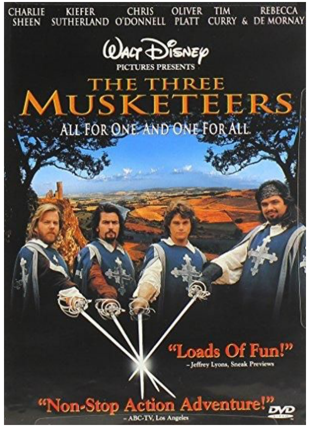
83.3 %	54.5 %

with reward bonus

91.3 %	84.2 %

Natural language is ambiguous

Correct queries yield wrong results (~4%)

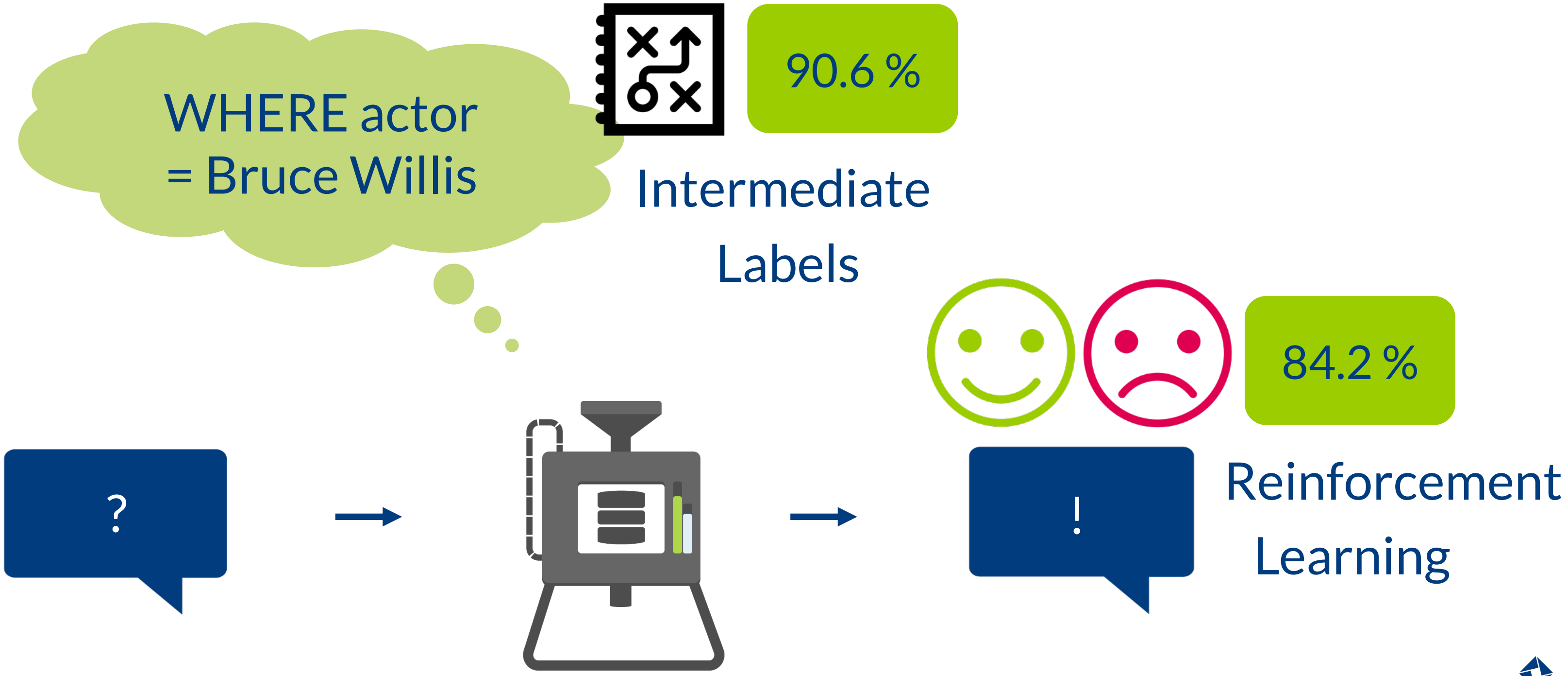


Images from Google Image Search

inovex

Comparison to supervised baseline

Intermediate labels provide a better feedback signal



Comparison to supervised baseline

Reinforcement learning requires a **LOT** of resources

	Reinforcement Learning	Intermediate Labels
96k	84.2 %	90.6 %
10k	19.3 %	89.1 %
96k	12h	1h

inovex

Comparison to soft lookup

Hard lookup achieves competitive results

Hard lookup		Soft lookup	
Reinforcement Learning	84.2 %	MemNN (ensemble)	83.5 %
Intermediate Labels	90.6 %	QA-system	90.7 %

inovex

Learnings

Conclusion

- We applied a Seq2Seq approach with pointer attention to create database queries from natural language questions.
- Our model achieves end-to-end trainability due to the usage of policy-based Reinforcement Learning and thereby avoids costly intermediate labels.
- Furthermore, we overcome local optima through exploration induced by count-based reward boni.

Outlook

- More complex questions & different corpora
- Improve sample-efficiency of RL
- Reduce latencies of database interaction

Read more

https://www.inovex.de/blog/

https://www.inovex.de/blog/seqpolicynet-nlp-elasticsearch/

http://www.aifb.kit.edu/images/d/d3/IAAI-19_paper_88.pdf

Thank you !

Sebastian Blank
Data Scientist

inovex GmbH
Ludwig-Erhard-Allee 6
76131 Karlsruhe

sebastian.blank@inovex.de