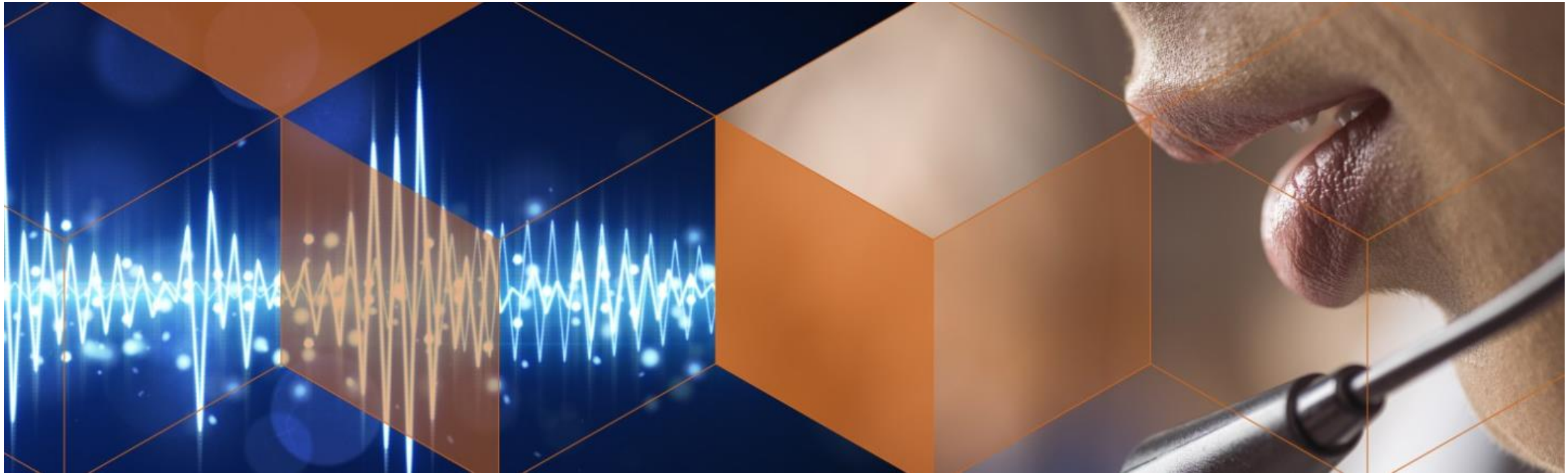
CONVERSATIONAL AI & KNOWLEDGE GRAPHS

LAMBDA Webinar

Prof. Dr. Jens Lehmann, Dr. Diego Collarana – Fraunhofer IAIS



Fraunhofer IAIS

LAMBDA - Learning, Applying, Multiplying Big Data Analytics

Twinning Coordination and Support Action, H2020-WIDESPREAD-2016-2017

Project Partners

- Institute Mihajlo Pupin, Serbia (Coordinator)
- Fraunhofer Institute for Intelligent Analysis and Information Systems, Germany
- Institute for Computer Science - University of Bonn, Germany
- Department of Computer Science - University of Oxford, UK

Primary Objectives

1. Strengthening the Human capital and Education, Research and Development capacities of “Mihajlo Pupin” Institute
2. Decreasing the existing European regional R&I disparity by fostering excellence in the Big Data Ecosystem areas,

Project ending on June 2021!!!

LAMBDA

AGENDA

- Motivation
- Conversational AI pipeline
- Knowledge Graphs for Conversational Assistants
- Question Answering over KGs (KGQA)
- Conclusions

Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS

Do more with data!

- Focus Artificial Intelligence/ Machine Learning
- 300 Data Scientists and AI Experts
- Basic Innovations, R&D, Industry Projects, Spinoffs

Image-Sources: www.flaticon.com, by Becris, Freepik

Fraunhofer IAIS

Fraunhofer IAIS Branch Dresden

- Focus on Conversational AI & Knowledge Graphs
- >15 colleagues (October 2020)
- „CEE-AI“ AI-Cluster established with TUD Rectorate
- Initial EU project, Industry projects & Industry workshops started
- Successful application for the AI Competence Centre Scads.AI

Quelle: https://www.transconnect-online.de

Why is Conversational AI important?

Conversational AI
Motivation

- Speech / Text is the most direct interface for humans (as long as thoughts cannot be read…)
- Increasing acceptance of conversational methods as universal interfaces
- Reductions of costs & complexity

Language is the most powerful instrument of cooperative action that humankind has, that any species we know of has.

(Randy Allen Harris, Voice Interaction Design)

How AI is Changing the Face of Customer Service

October 15, 2018

How AI is Changing the Face of Customer Service

October 15, 2018

23,687 views | Jun 1, 2018, 12:28am

The Digital Transformation Of Accounting And Finance – Artificial Intelligence, Robots And Chatbots

How AI-powered chatbots are unlocking business value today

Chatbots can streamline engagement between consumers and brands to improve customer service experience

14 NOVEMBER 2018 - 12:49 *by* RUDEON SNELL

HARVARD BUSINESS REVIEW

Collaborative Intelligence: Humans and AI Are Joining Forces

by H. James Wilson and Paul R. Daugherty

FROM THE JULY–AUGUST 2018 ISSUE

The Future Of Medical Technology: Chatbots

Fraunhofer IAIS

Chatbot global market
Motivation

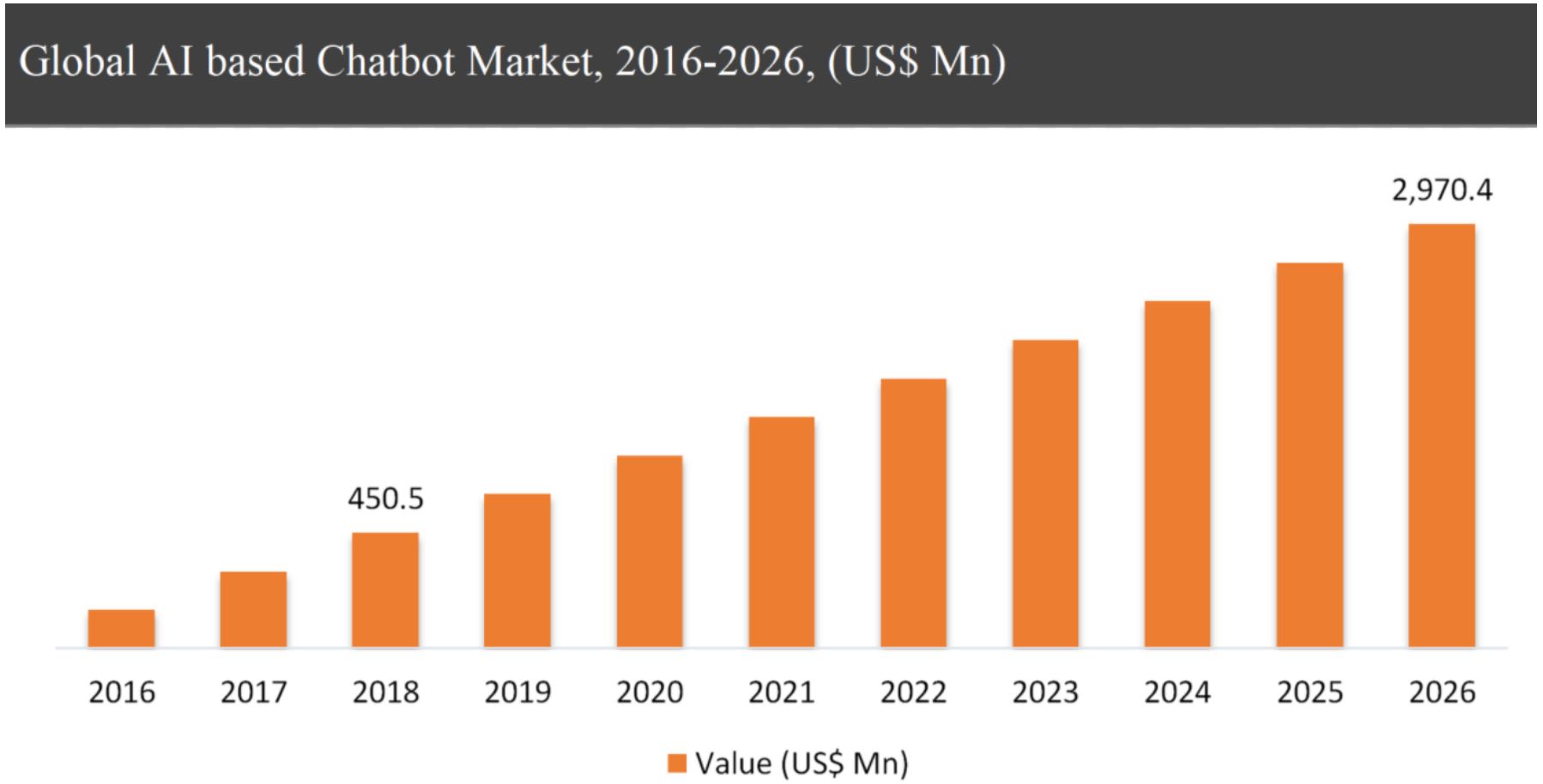
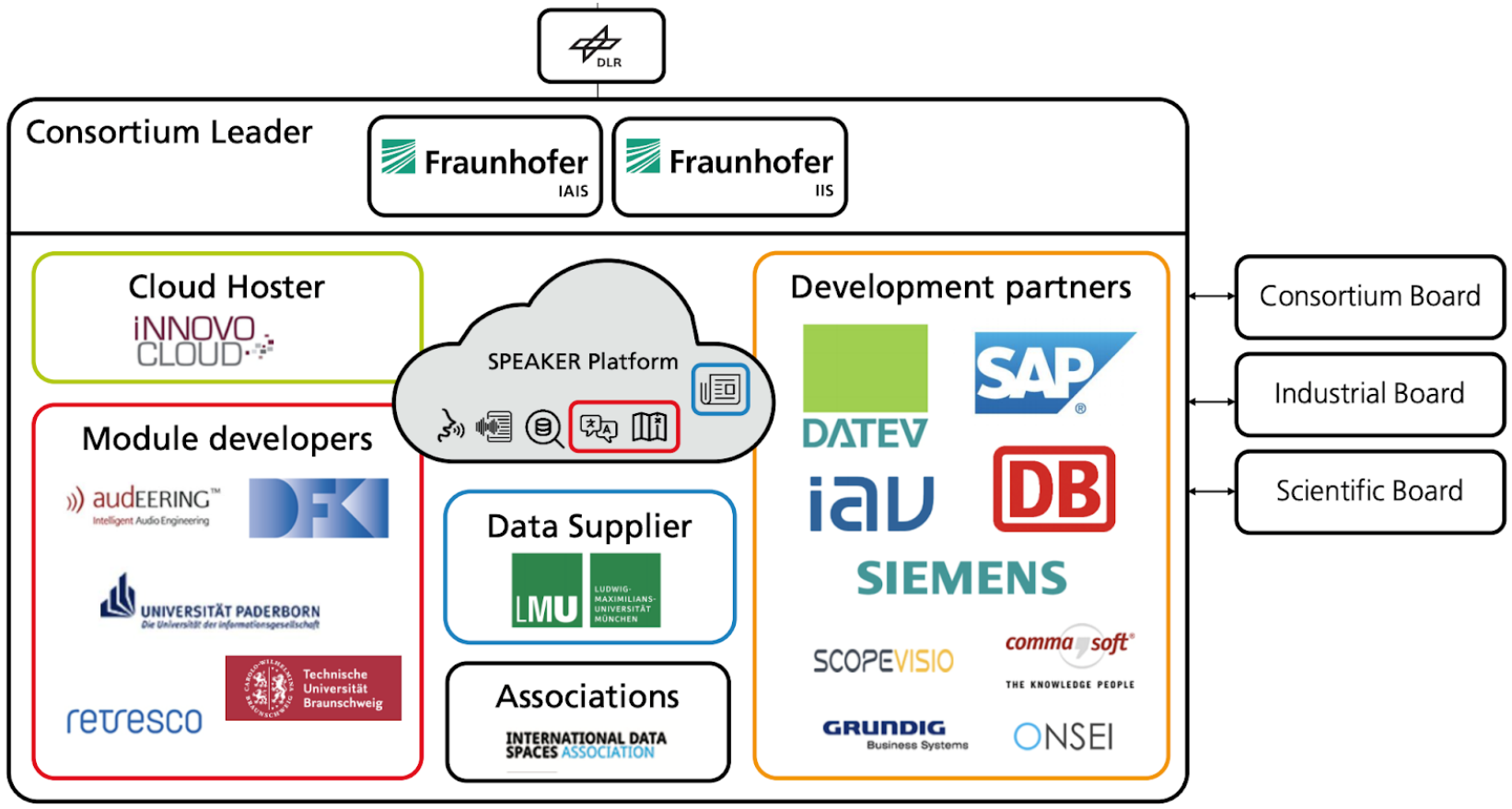


Image taken from https://www.alltheresearch.com/report/223/ai-based-chatbot-market-market

SPEAKER project
Industrial Speech Assistant Platform

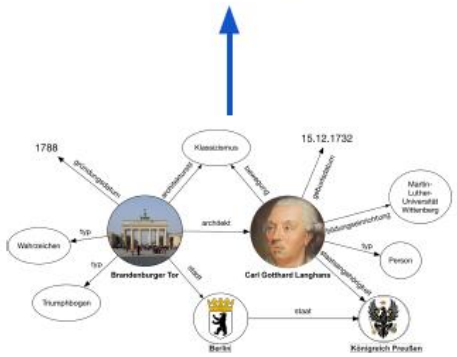
- Conversational AI
- Question Answering
- Knowledge Graphs
- Speech
- Privacy
- Customizable & Domain-specific



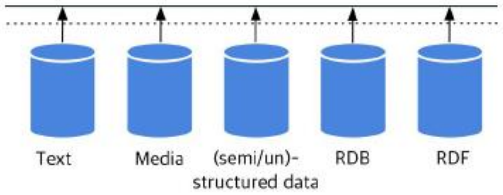
Conversational AI

Conversational Assistants
Background

We build conversational AI platforms



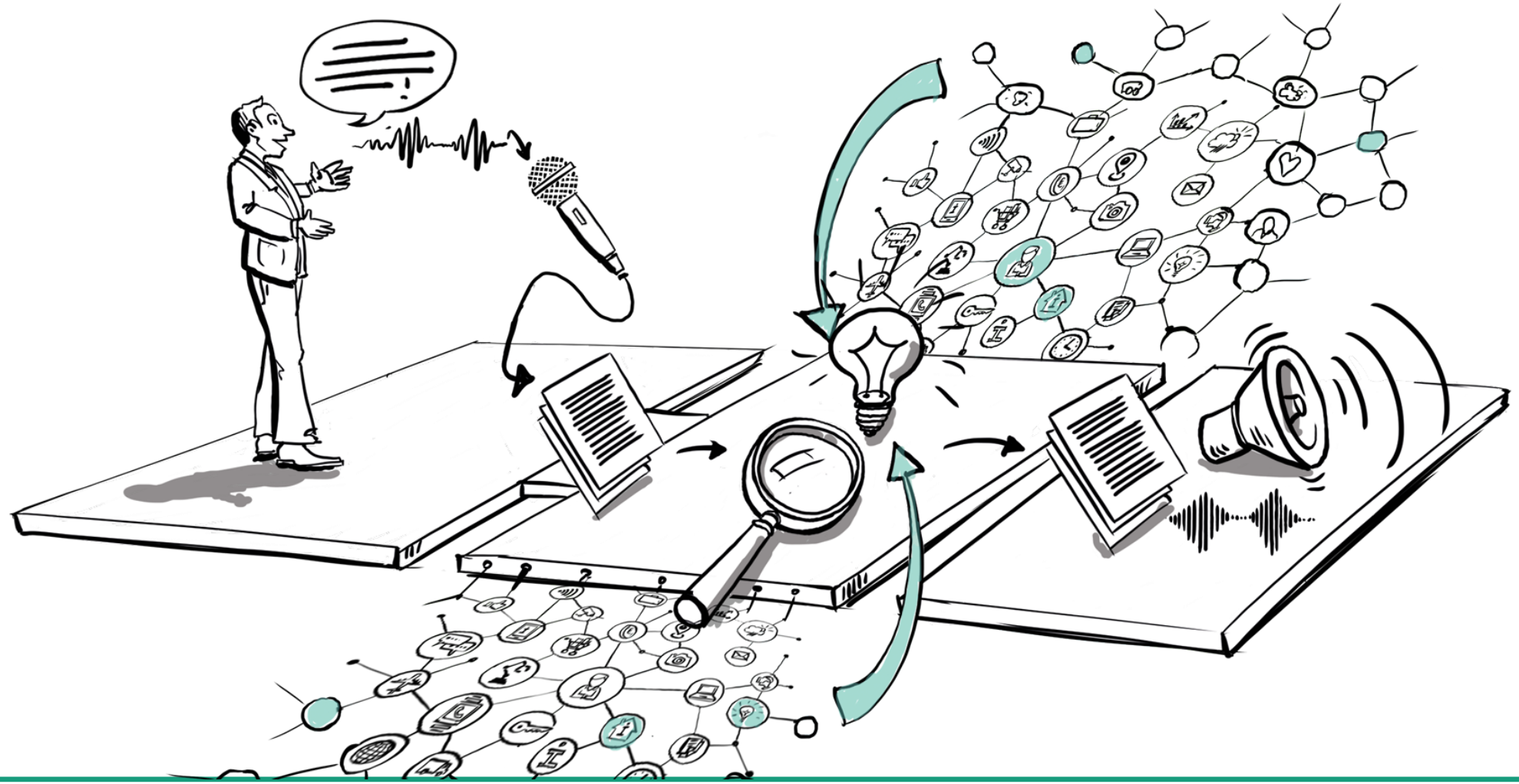
Powered by knowledge graphs



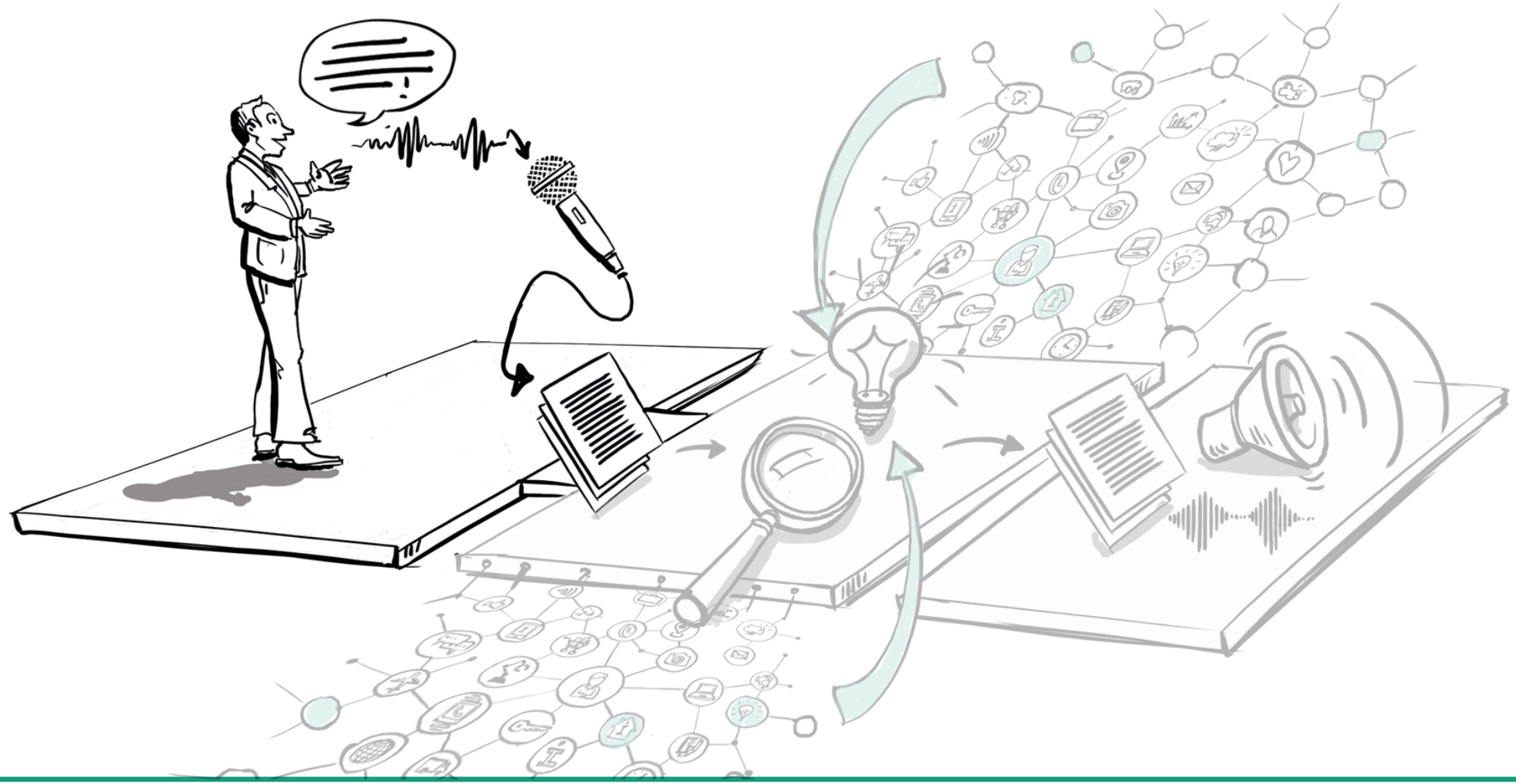
Obtained by integrating heterogeneous data

Source: Michael Galkin https://migalkin.github.io/talks/2019-11-16-moscownlp

Conversational Assistant
Pipeline

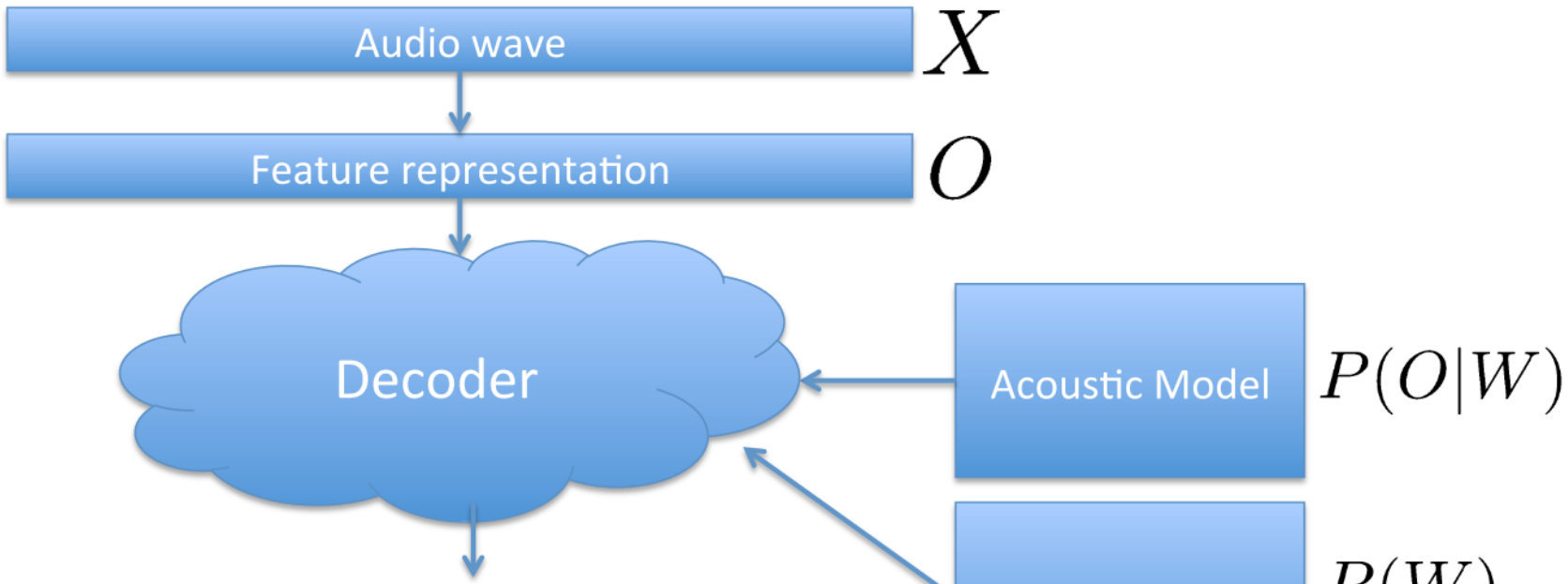


Automatic Speech Recognition (ASR)



Traditional ASR pipeline

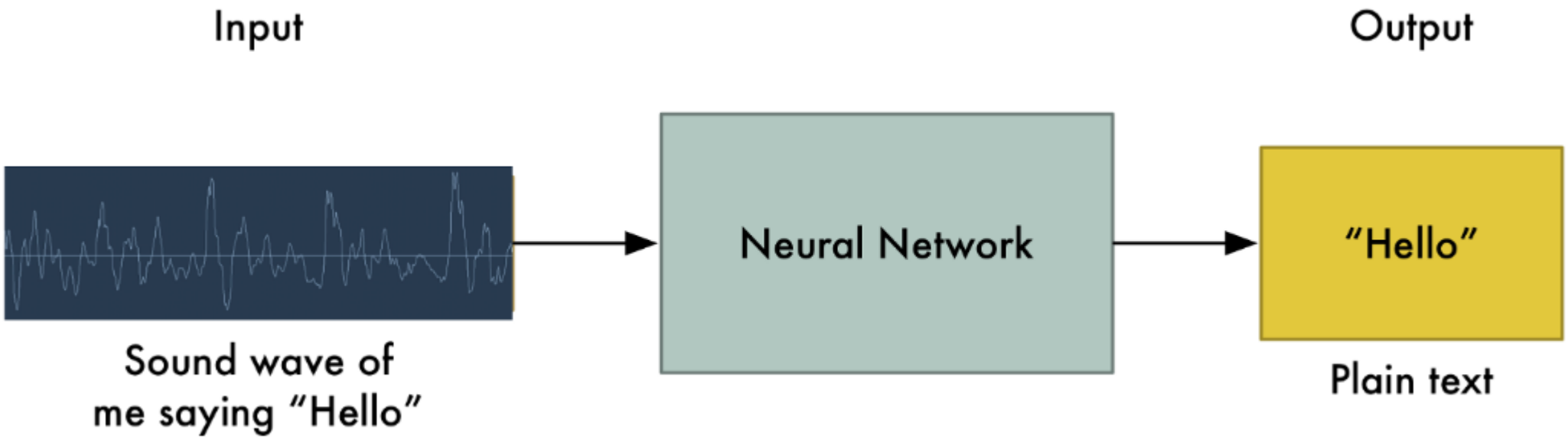
Break problem into several key components



Gales & Young, 2008
Jurafsky & Martin, 2000

Source: Adam Coates http://www.apcoates.com/

End-to-End ASR approach
Using neuronal networks



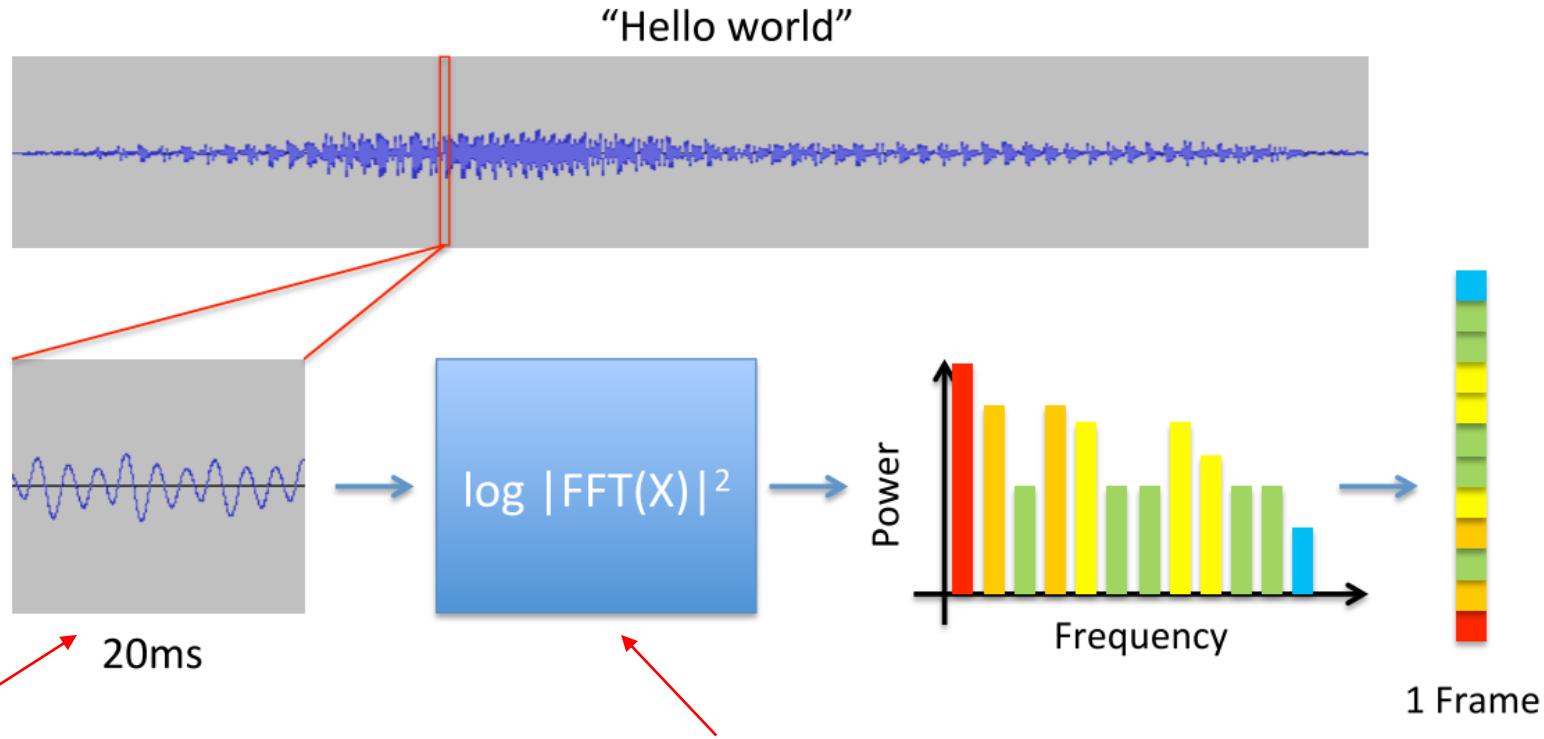
Supervised Learning

We will give many examples for the neural network to learn, a tuple (audio, text)

Source: Adam Coates http://www.apcoates.com/

End-to-End ASR approach

Audio pre-processing

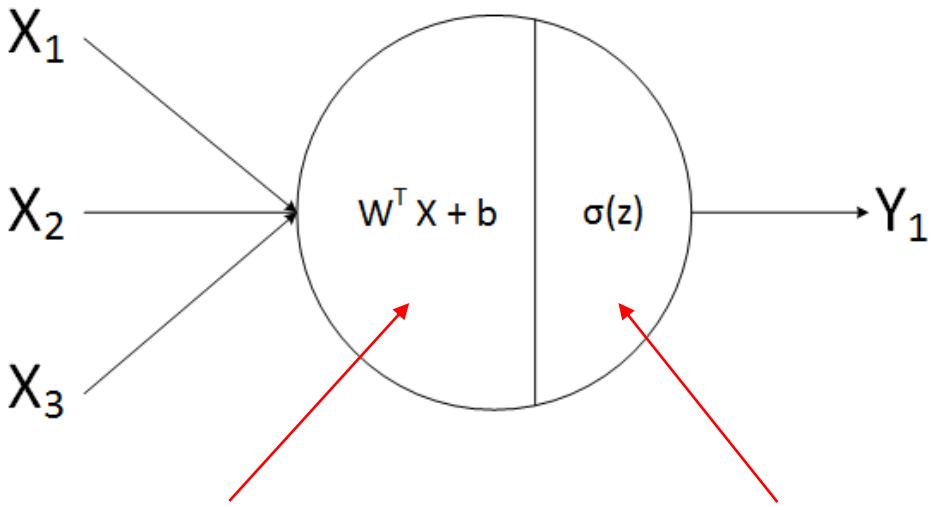


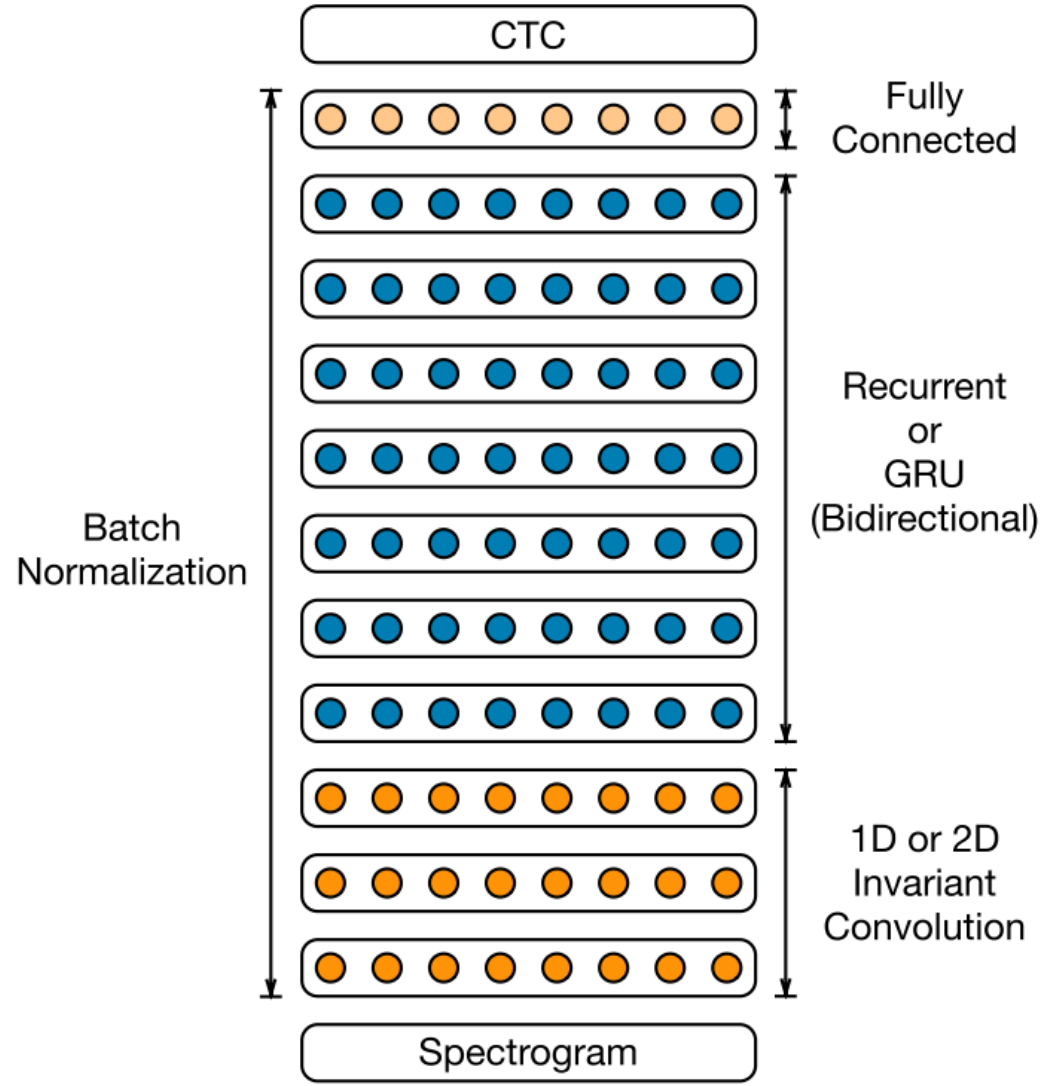
Take a small window (e.g., 20ms) of waveform.

Compute FFT and take magnitude. (i.e., power)

Source: Adam Coates http://www.apcoates.com/

End-to-End ASR approach
DeepSpeech2 model

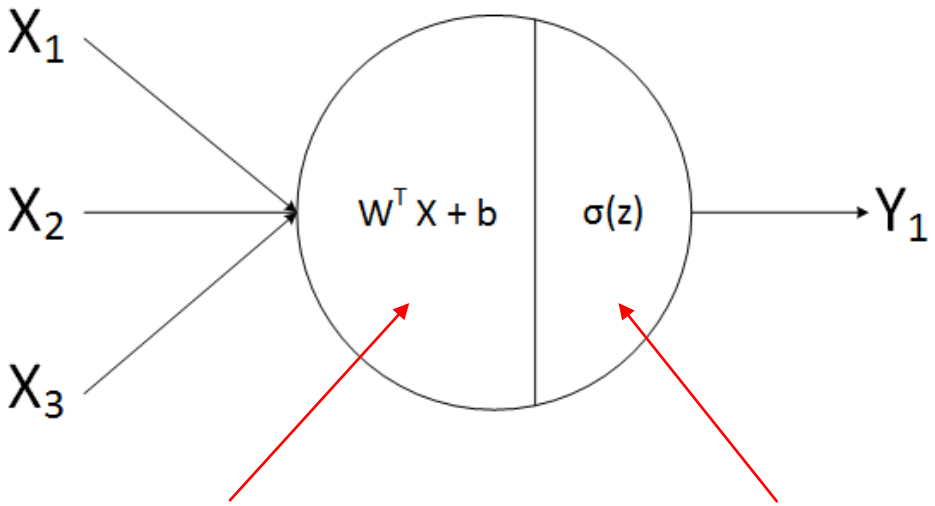




Model DeepSpeech 2: https://arxiv.org/abs/1512.02595

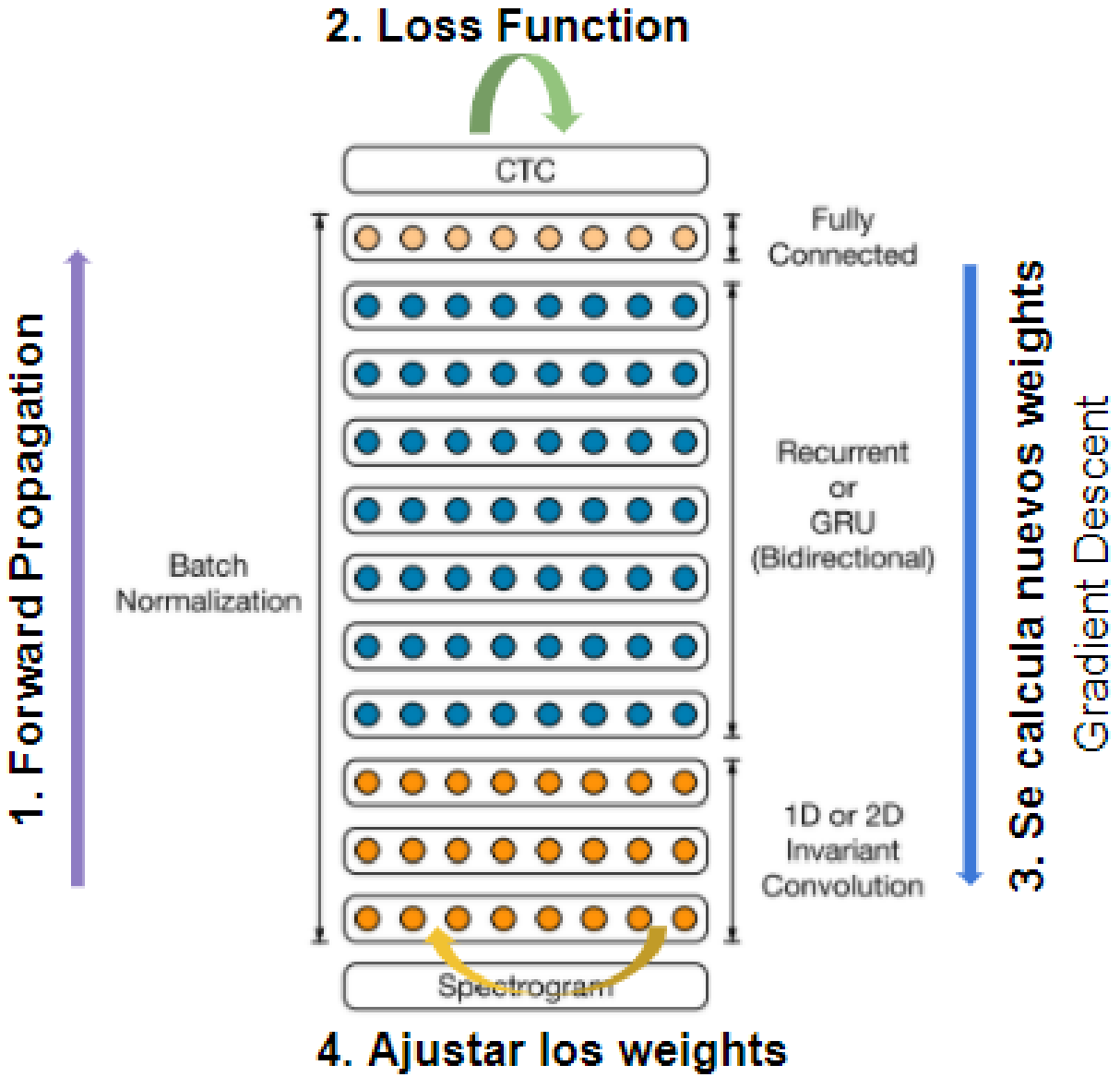
End-to-End ASR approach

DeepSpeech2 model



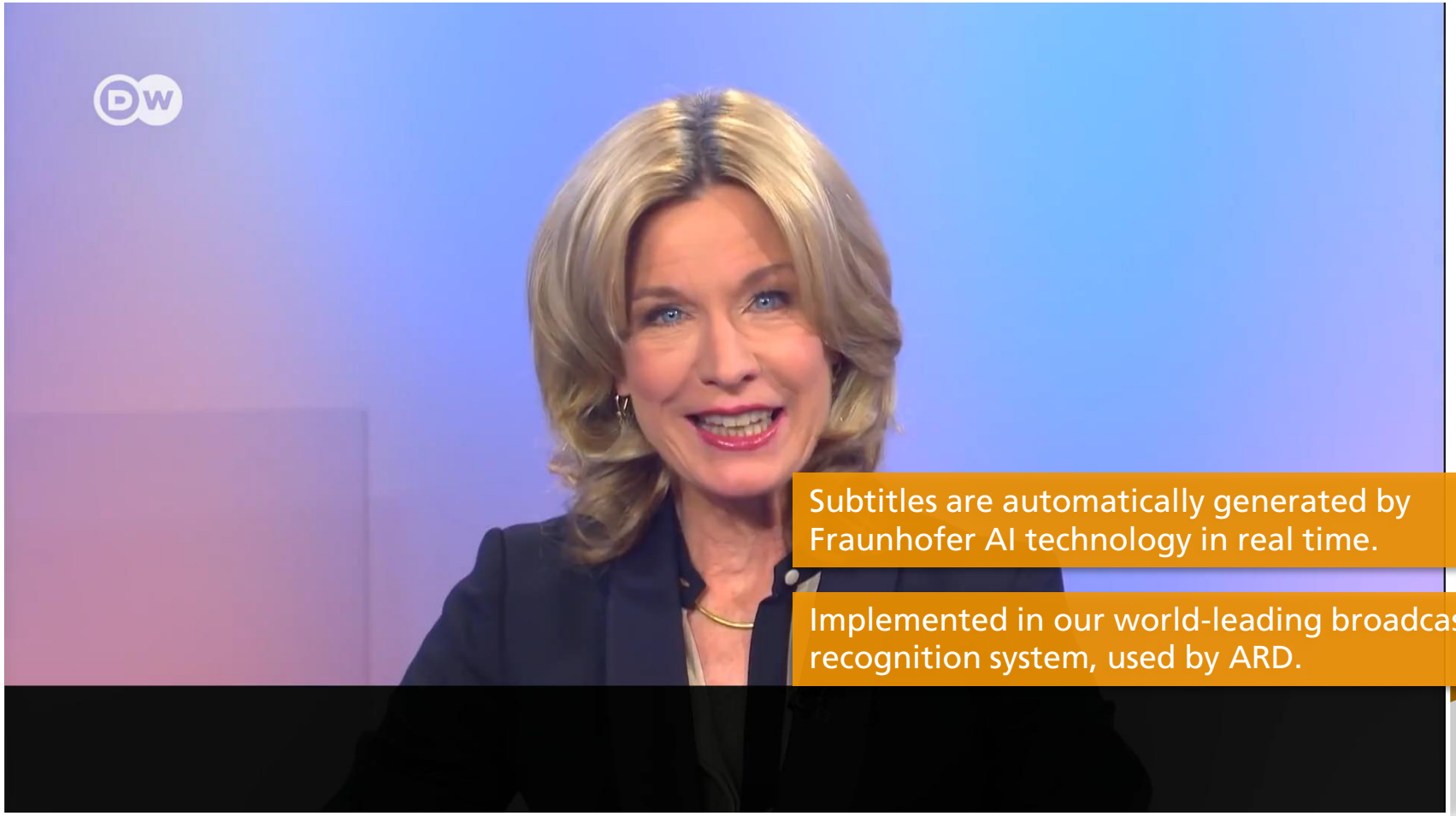
Linear function with weights (W) and biases (b)

Activation function



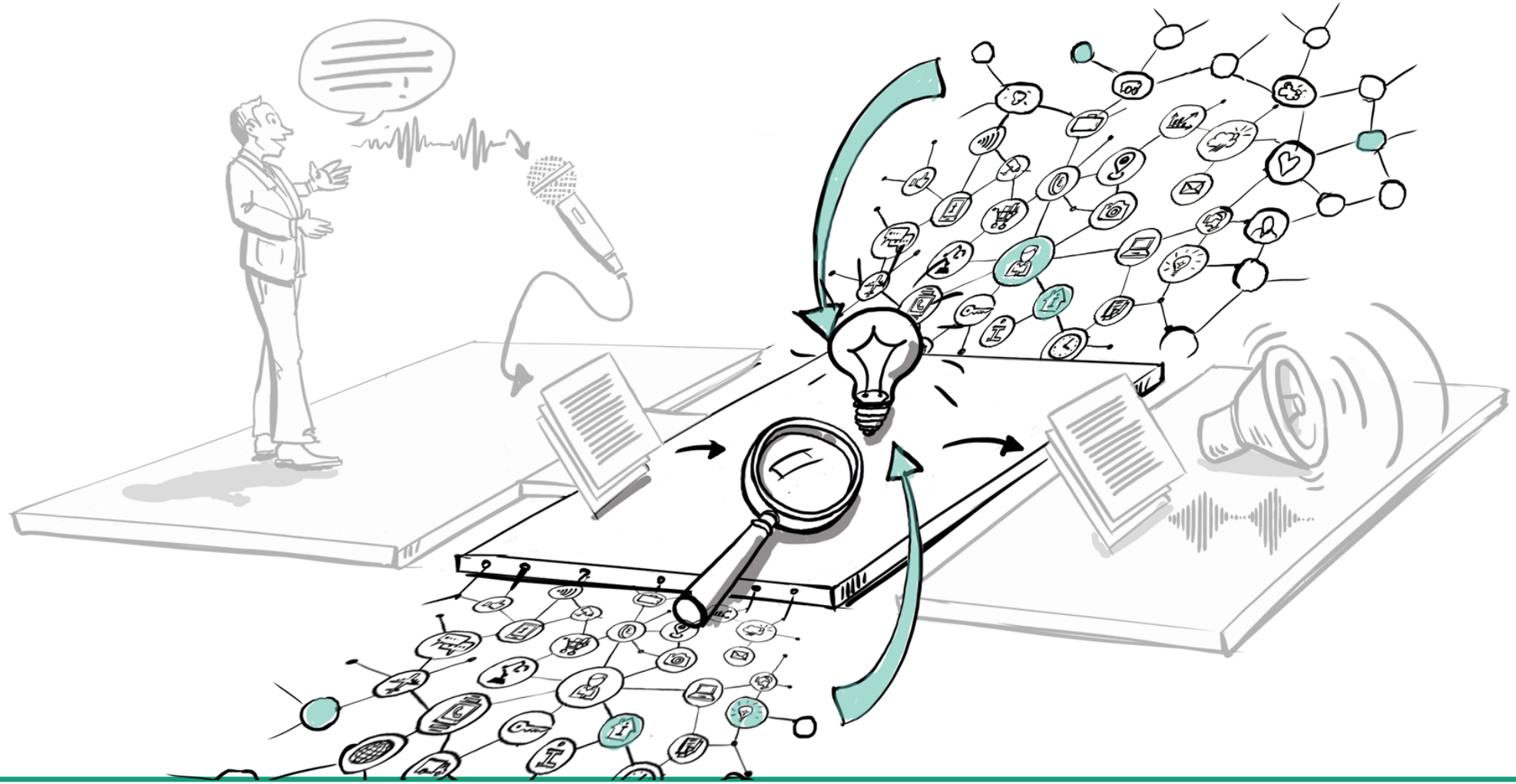
Model DeepSpeech 2: https://arxiv.org/abs/1512.02595

Speech Recognition Example



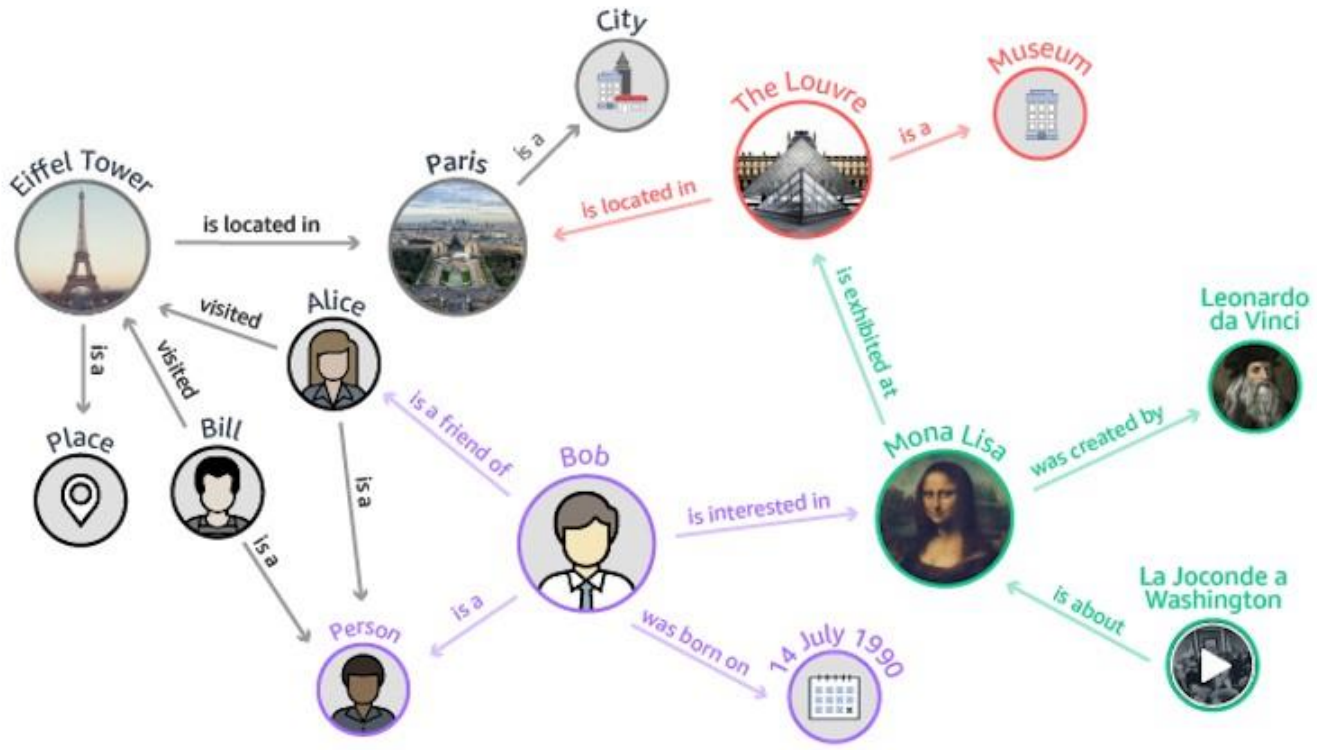
Speech Recognition: Saxony Parliament

Core Dialogue System – Powered by Knowledge Graphs



What is a knowledge graph?
Definition

"A knowledge graph is a large network of entities, their semantic types plus properties, and relationships between those entities."



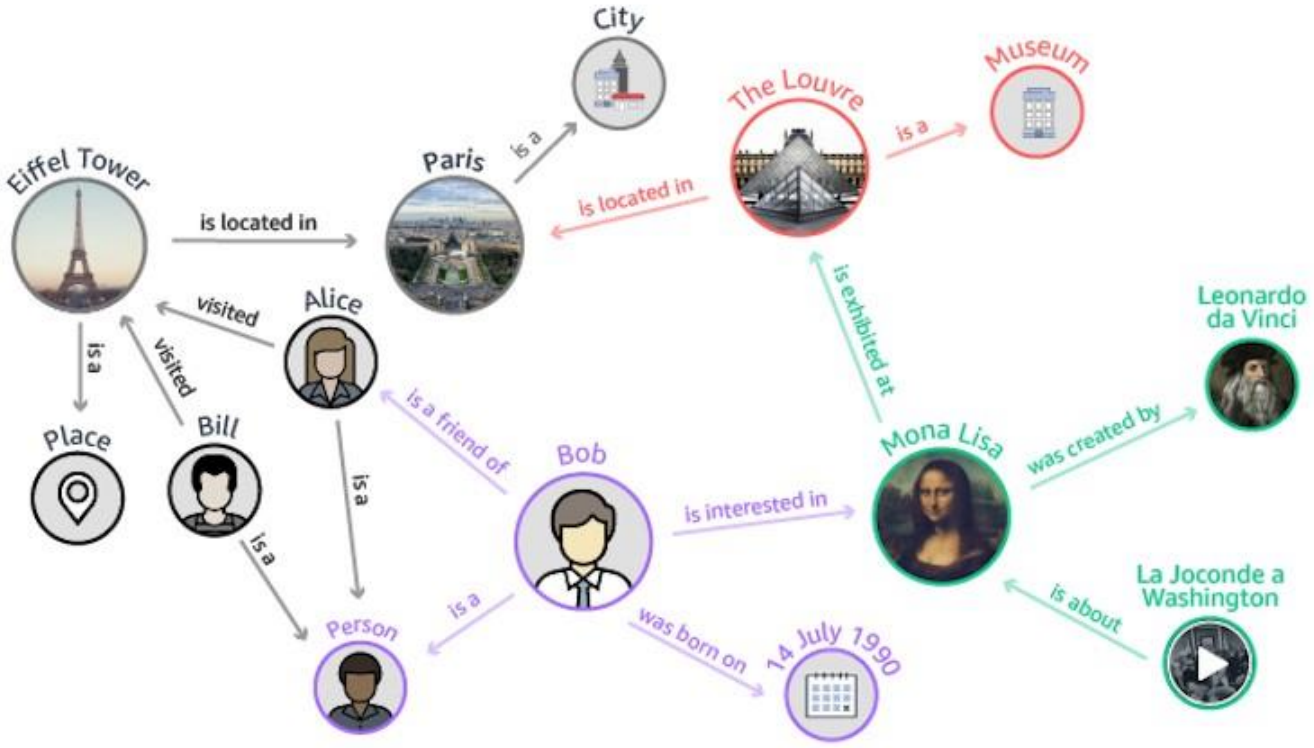
What is a knowledge graph?

Technically we use RDF a family of W3C specifications

RDF is a generic data modeling language based on the idea of making statements about the things in the form of:

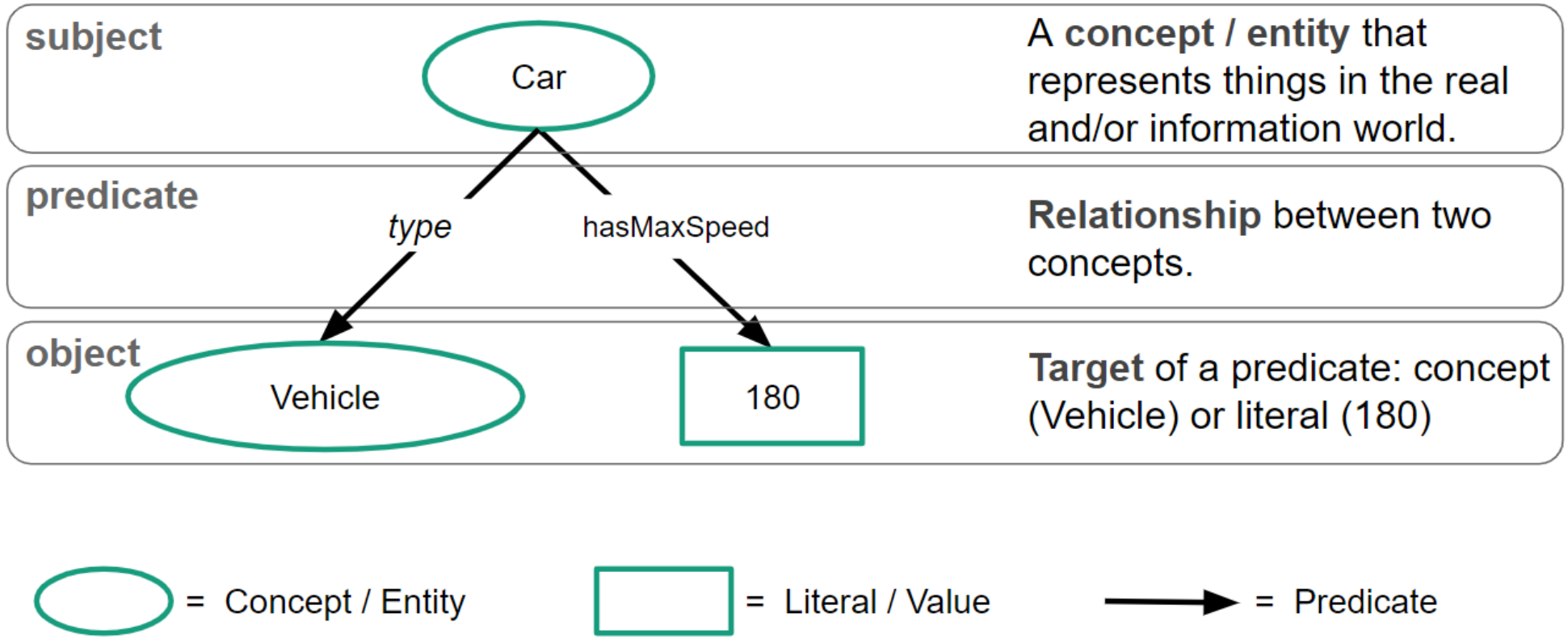
<subject> <predicate> <object>

which are known as “RDF triples”.



Fraunhofer IAIS

What is a knowledge graph?
Technically we use RDF a family of W3C specifications



What is a knowledge graph?

Technically we use RDF a family of W3C specifications

ex:EiffelTower a ex:Place .

ex:EiffelTower ex:locatedIn ex:Paris .

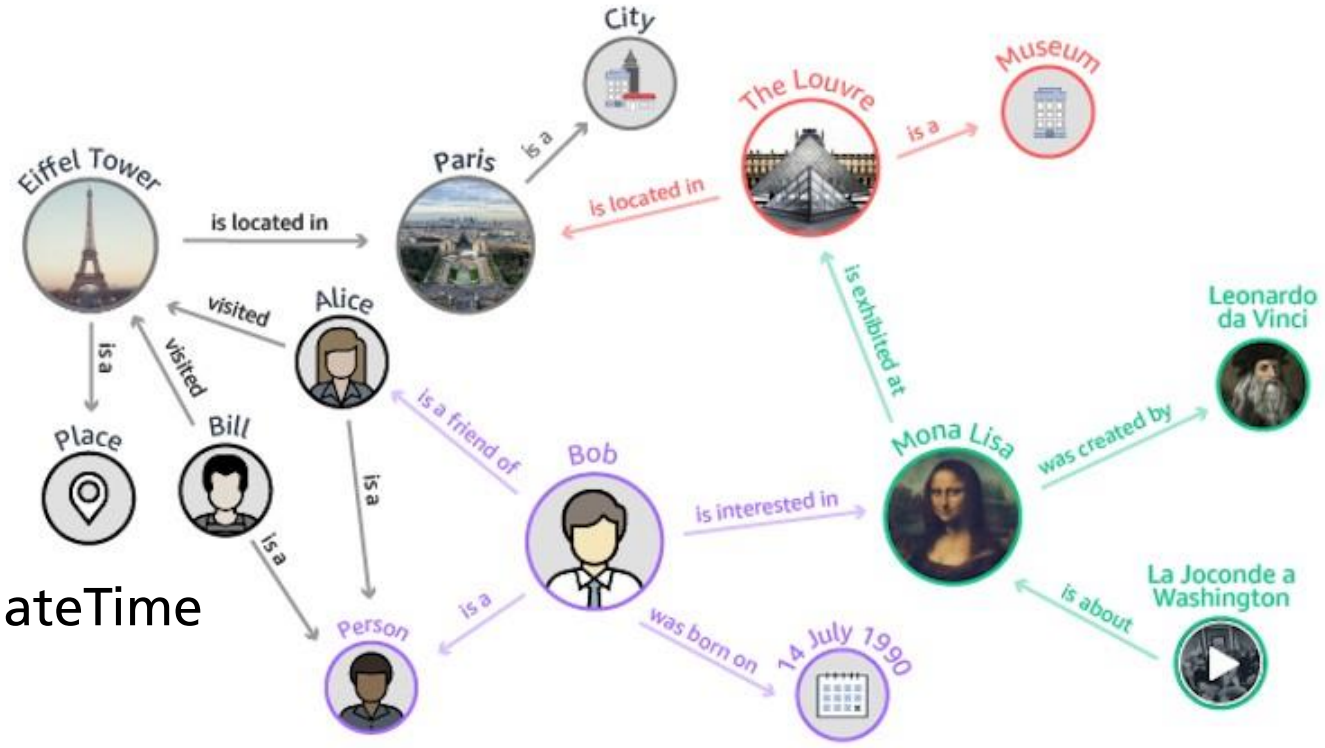
ex:Paris a ex:City .

ex:Alice ex:visited ex:EiffelTower .

ex:Bob ex:friendOf ex:Alice .

ex:Bob ex:bornOn “14.07.1990”^^xsd:dateTime

…



Serialization formats: Turtle, **N-Triples**, JSON-LD, N-Quads, RDF/XML

Fraunhofer IAIS

What is a knowledge graph?

Technically we use RDF a family of W3C specifications

ex:EiffelTower a ex:Place .

ex:EiffelTower ex:locatedIn ex:Paris .

ex:Paris a ex:City .

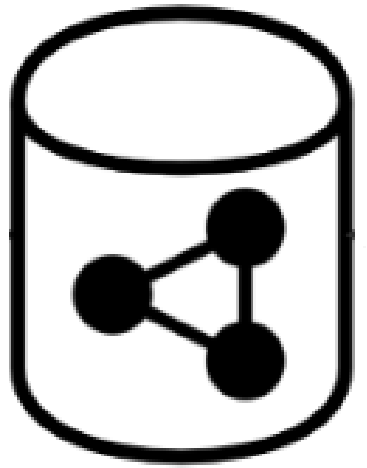
ex:Alice ex:visited ex:EiffelTower .

ex:Bob ex:friendOf ex:Alice .

ex:Bob ex:bornOn “14.07.1990”^^xsd:dateTime

…

Triple store



Virtuoso

Stardog

Amazon Neptune

…

Serialization formats: Turtle, **N-Triples**, JSON-LD, N-Quads, RDF/XML

What is a knowledge graph?

Technically we use RDF a family of W3C specifications

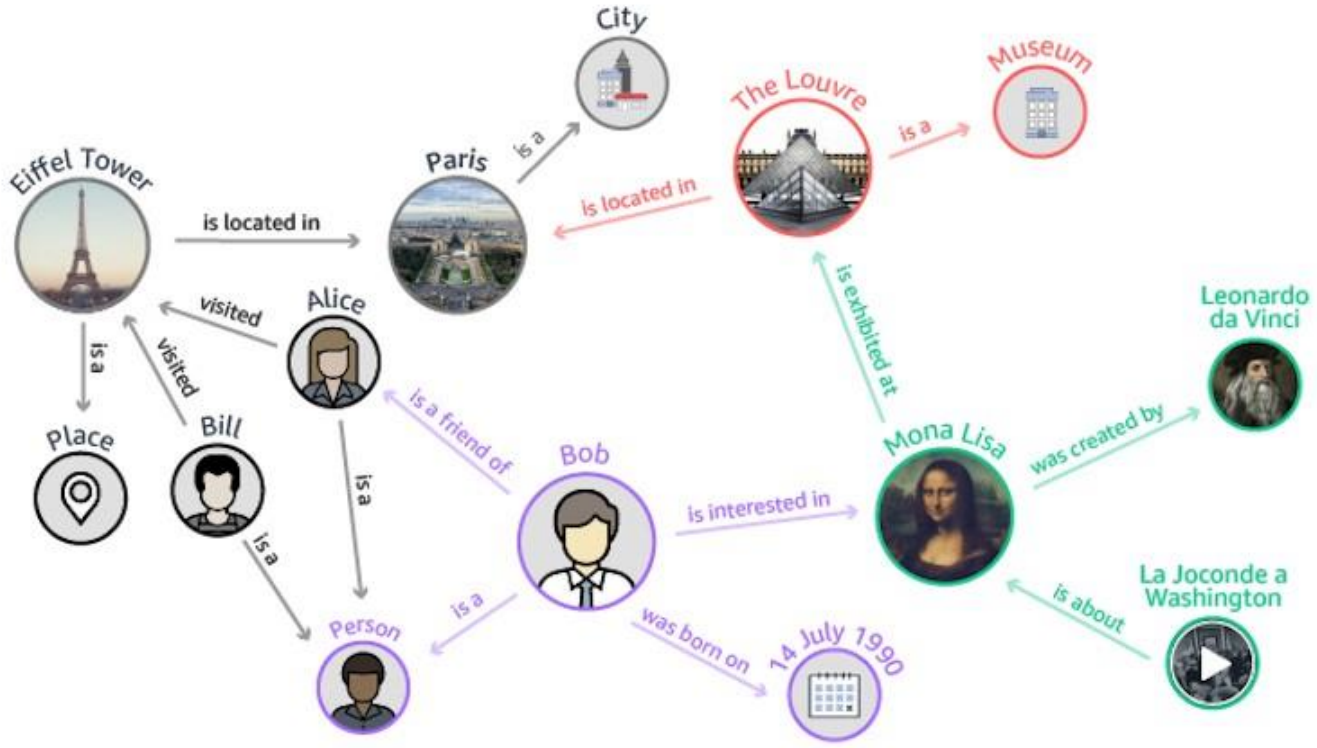
Who **visited** the **Eiffel Tower**?

↓

```
SELECT ?uri WHERE {
    ?uri ex:visited ex:EiffelTower.
}
```

↓

ex:Alice, ex:Bill



SPARQL query language

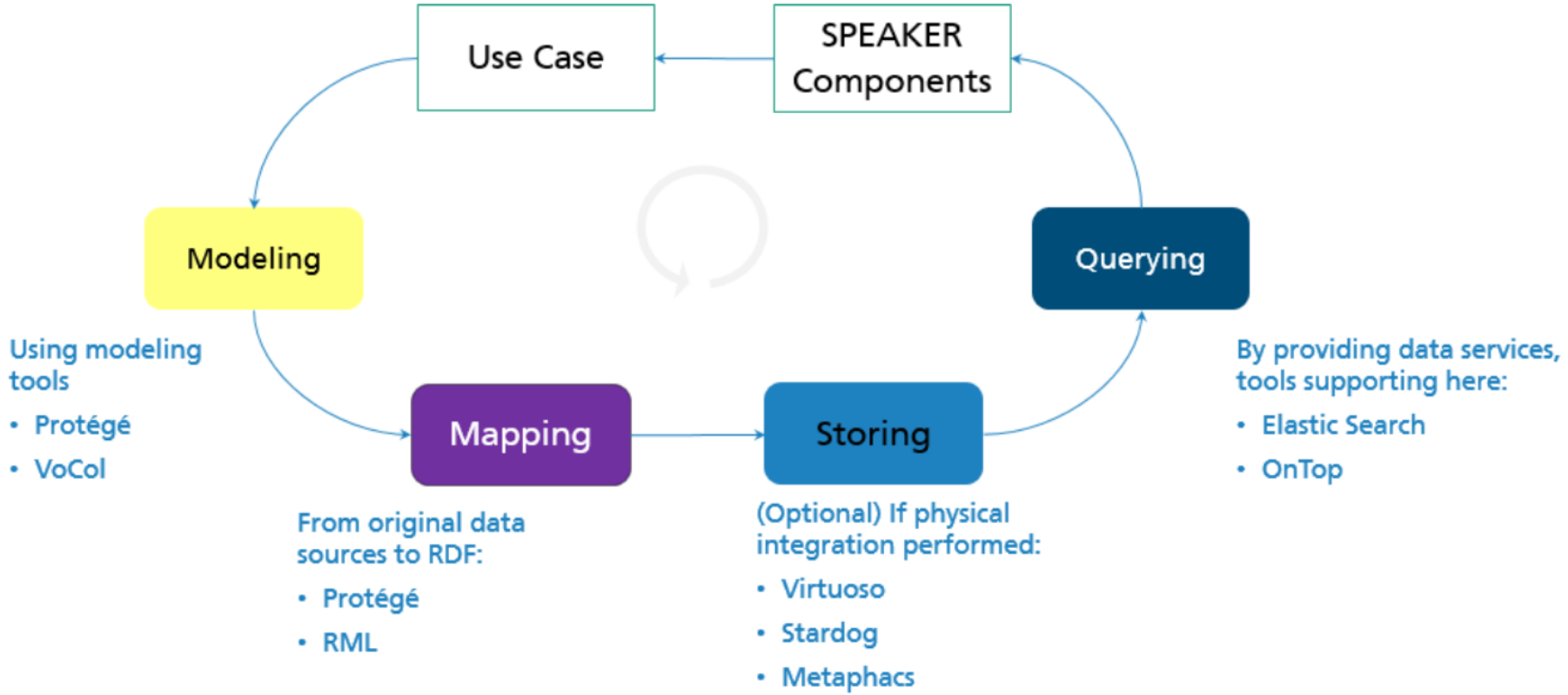
Fraunhofer IAIS

How to build a knowledge graph from enterprise data?

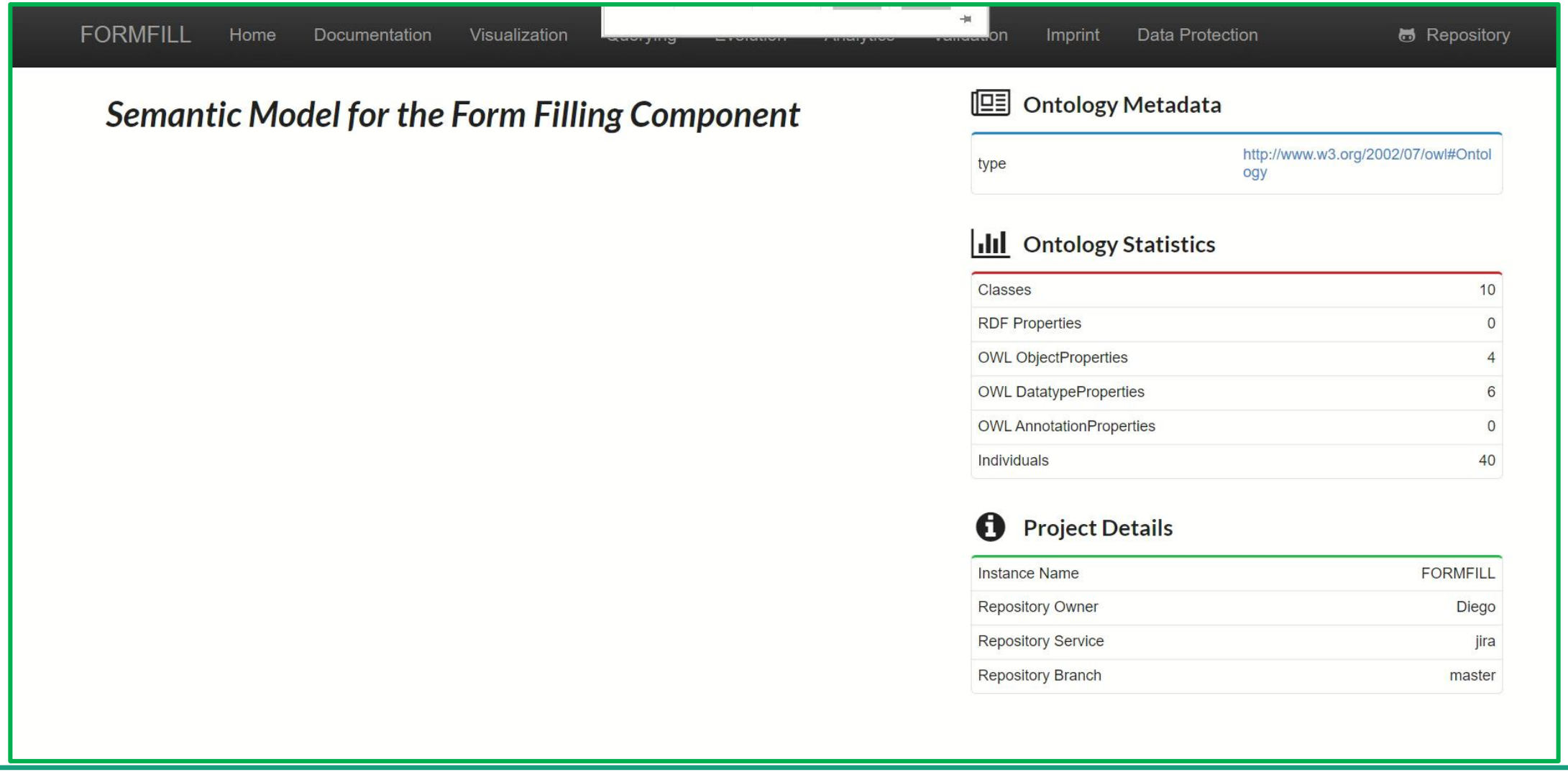
How to build a knowledge graph?
Is it expensive?

How to build a knowledge graph?
Smart Data Connector, easy and enjoyable

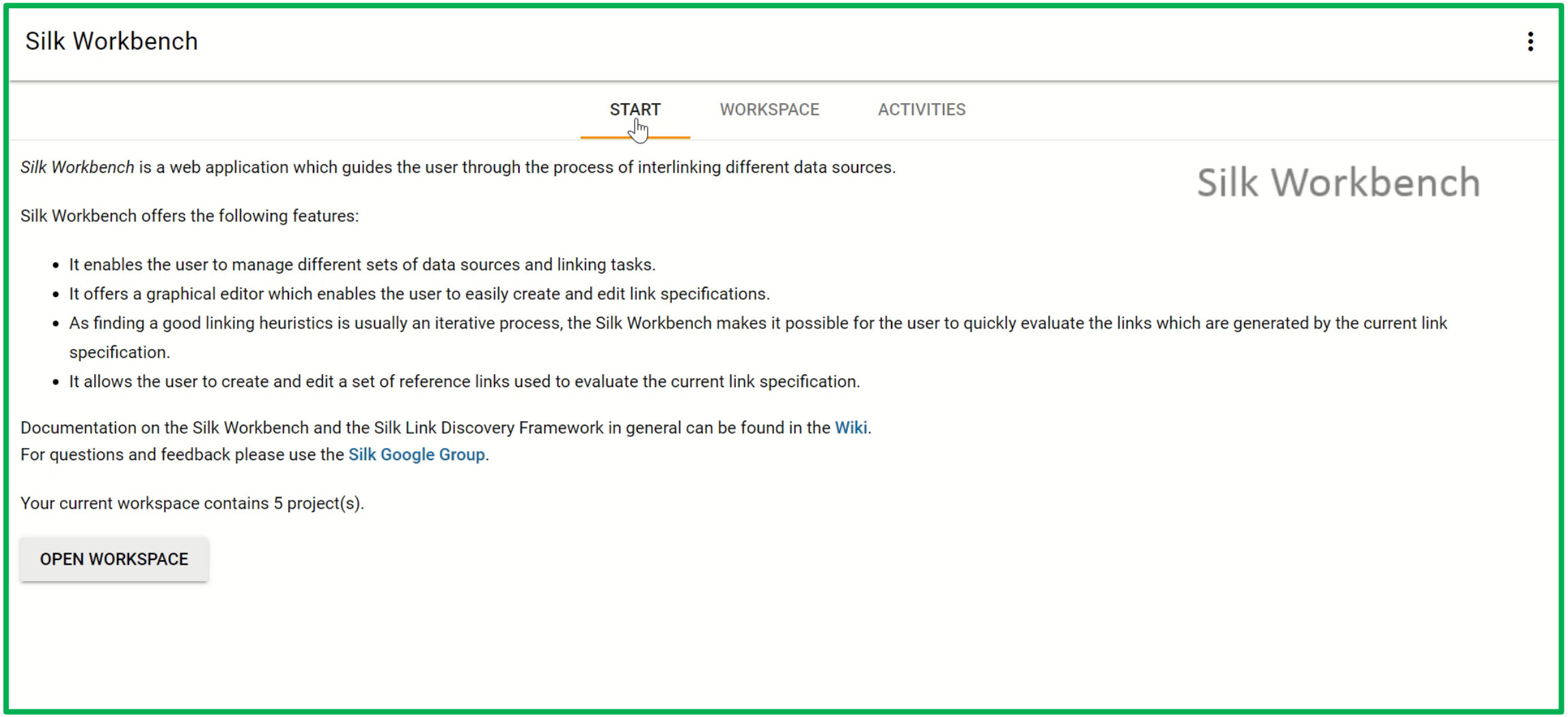
The Smart Data Connector (**SDC**) is a solution to connect heterogeneous enterprise data into actionable knowledge, with the motto **model** and **connect**



How to build a knowledge graph?
Modeling the domain



How to build a knowledge graph?
Mapping and connect your data



How to build a knowledge graph?
Smart Data Connector, easy and enjoyable

MVP Features

Connectors to SQL, CSV, JSON, XML

Provide the means to create mappings and transformers from SQL table, CSV and JSON files to RDF.

Query End-Point

Provide a SPARQL end-point, to query the connected data.

Entity Index

Provide a rich API, which allows for a quick keyword search on business entities from the Connectors.

Metadata Index

Indexing as well metadata, both customer provided and open, e.g., **full medical thesaurus.** By indexing not dummy JSON objects but semantic enriched JSON-LD objects.

Smart Data Connector 0.0.1 OAS3

/openapi.json

API for the SPEAKER Smart Data Connector

Elastic Search Index

GET **/entities** Entities

GET **/relations** Relations

GET **/search** Search

Query Knowledge Graph

POST **/query** Query

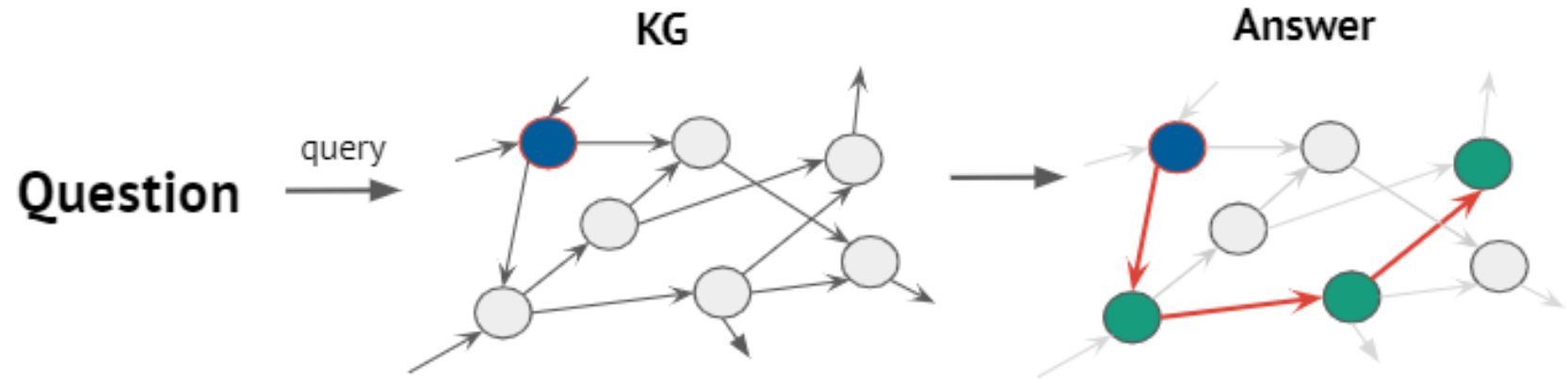
Data Services

Knowledge Graph-based Question Answering

Knowledge Graph-based Question Answering
How it works

KGQA

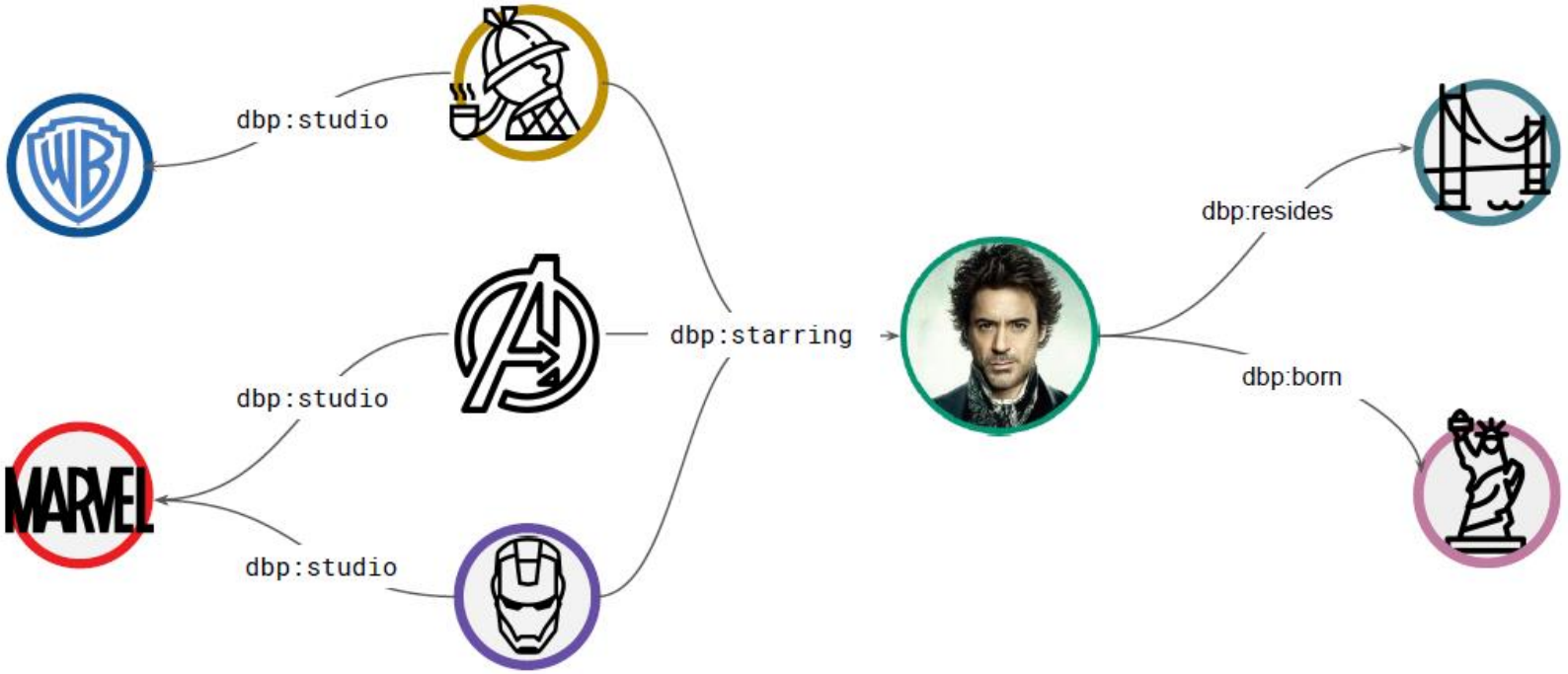
Knowledge Graph-based
Question Answering



Source: Michael Galkin https://migalkin.github.io/talks/2020-12-04-kgqa

Knowledge Graph-based Question Answering

How it works



Q: How many Marvel movies was Robert Downey Jr. casted in?

Knowledge Graph-based Question Answering

How it works

KGQA

How many **Marvel** **movies** was **Robert Downey Jr.** **casted** in?

```
SELECT COUNT(?uri) WHERE {
    ?uri dbp:studio dbr:Marvel_Studios.
    ?uri dbo:starring dbr:Robert_Downey_Jr
}
```

Source: Michael Galkin https://migalkin.github.io/talks/2020-12-04-kgqa

Knowledge Graph-based Question Answering

How it works

KGQA

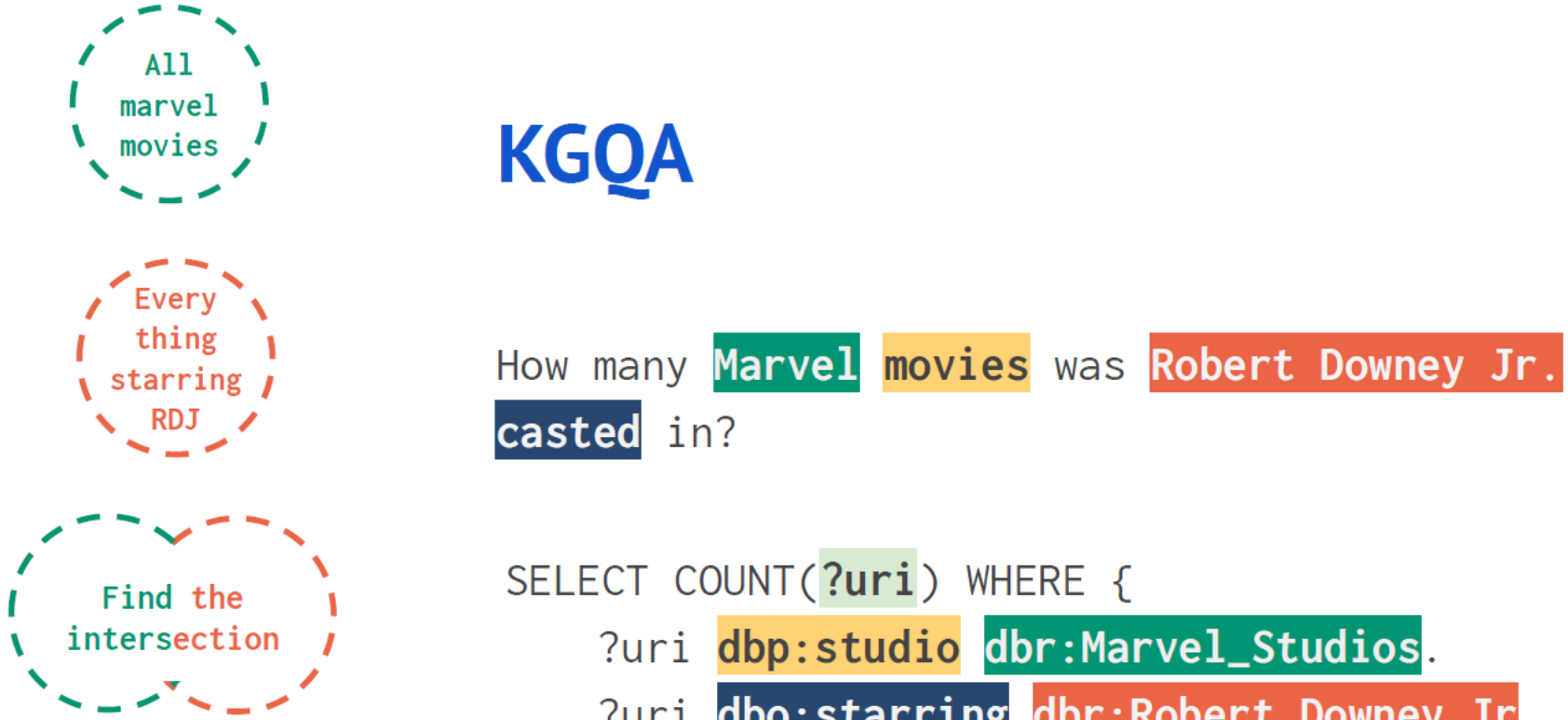
How many Marvel movies was Robert Downey Jr. casted in?

```
SELECT COUNT(?uri) WHERE {
    ?uri dbp:studio dbr:Marvel_Studios.
    ?uri dbo:starring dbr:Robert_Downey_Jr
}
```

Source: Michael Galkin https://migalkin.github.io/talks/2020-12-04-kgqa

Knowledge Graph-based Question Answering

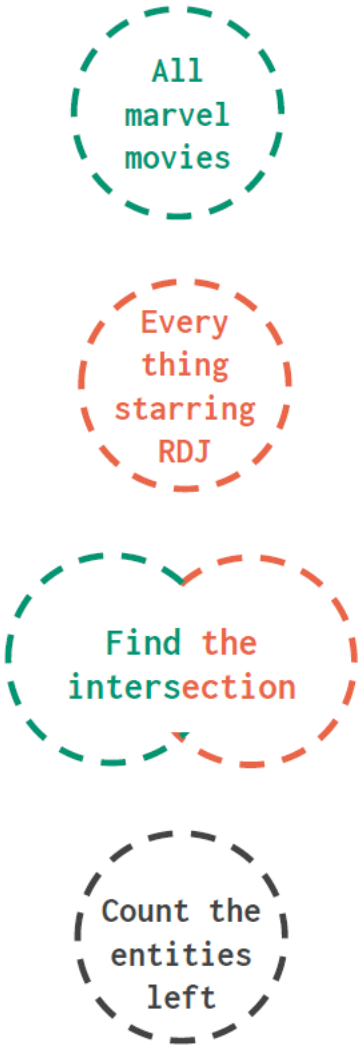
How it works



Source: Michael Galkin https://migalkin.github.io/talks/2020-12-04-kgqa

Knowledge Graph-based Question Answering

How it works



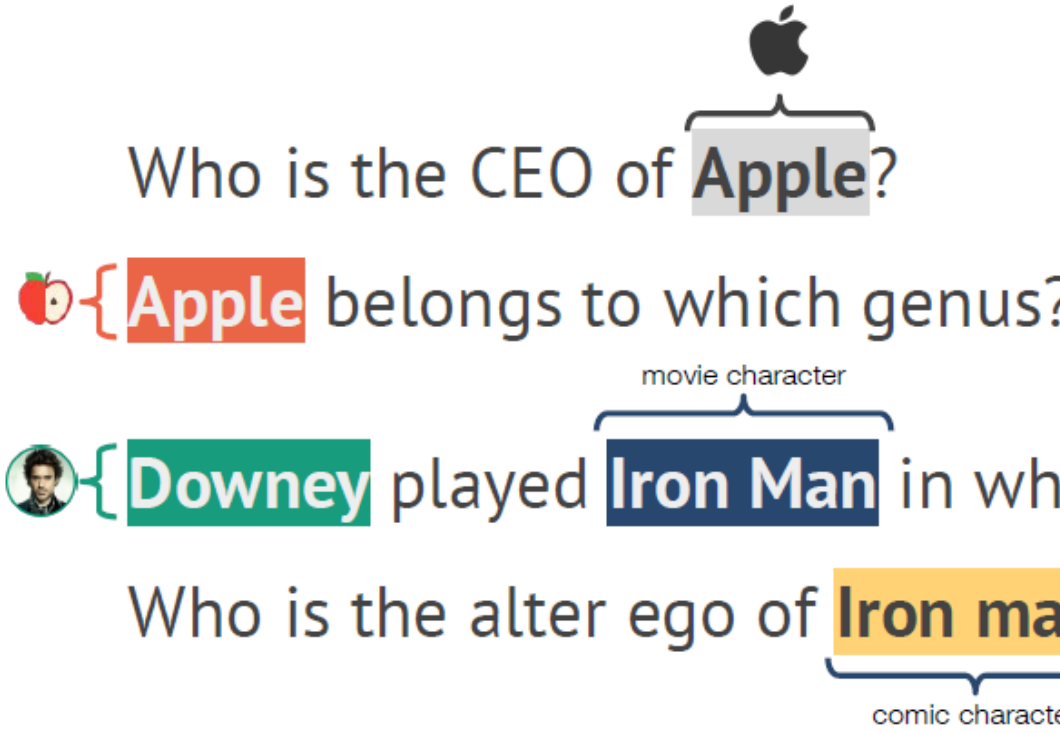
KGQA

How many **Marvel** **movies** was **Robert Downey Jr.** **casted** in?

```
SELECT COUNT(?uri) WHERE {
    ?uri dbp:studio dbr:Marvel_Studios.
    ?uri dbo:starring dbr:Robert_Downey_Jr
}
```

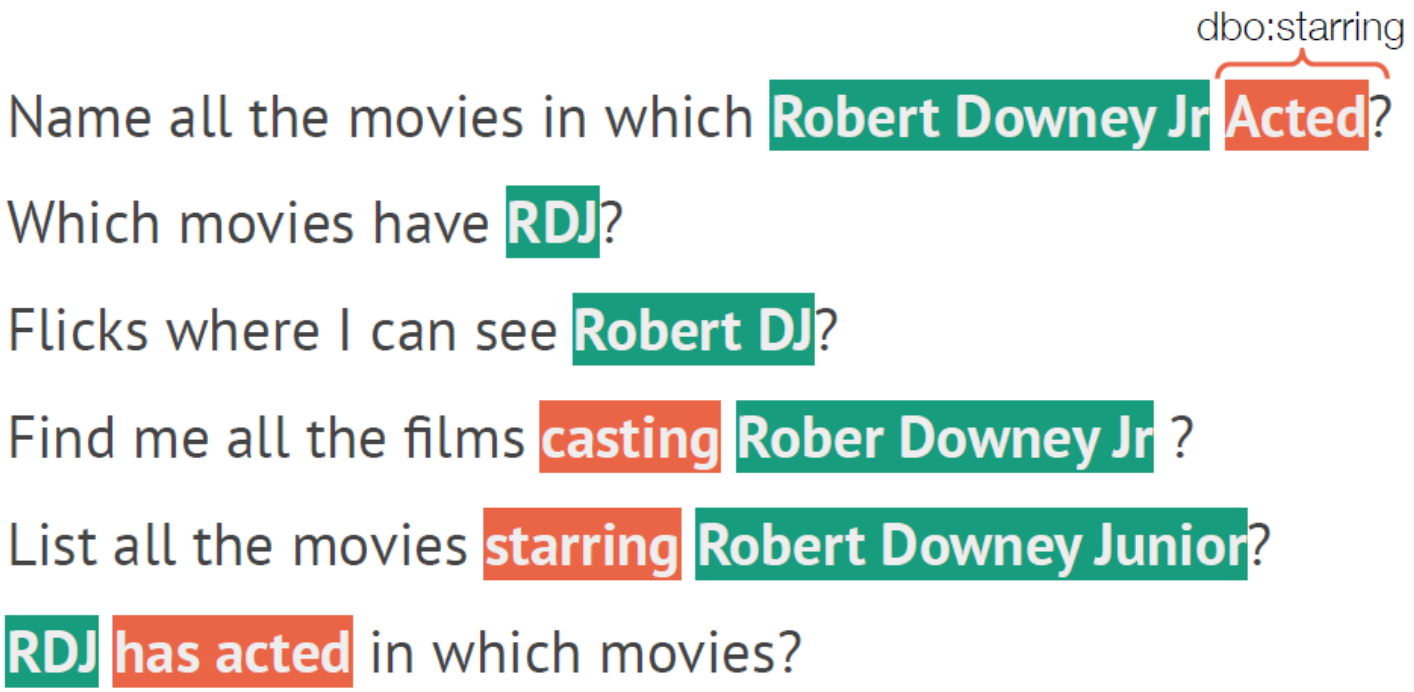
Source: Michael Galkin https://migalkin.github.io/talks/2020-12-04-kgqa

Knowledge Graph-based Question Answering
Entity Linking



Source: Michael Galkin https://migalkin.github.io/talks/2020-12-04-kgqa

Knowledge Graph-based Question Answering
Relation Linking



Source: Michael Galkin https://migalkin.github.io/talks/2020-12-04-kgqa

Knowledge Graph-based Question Answering

Pre-defined SPARQL templates

Natural language

How many {Entity} did {Entity} star in?

Pre-defined template

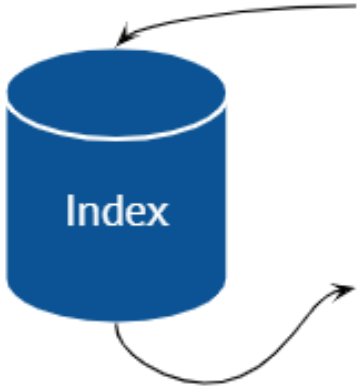
```
SELECT COUNT(?x) WHERE {
    ?x rdf:type ?c.
    ?x wdt:P161 ?y
}
```

Source: Michael Galkin https://migalkin.github.io/talks/2020-12-04-kgqa

Knowledge Graph-based Question Answering
Pre-defined SPARQL templates

How many {Entity} did {Entity} star in?

How many films did Leonardo DiCaprio star in?




```
SELECT COUNT(?x) WHERE {
    ?x rdf:type wd:Q11424.
    ?x wdt:P161 wd:Q38111
}
```

wd:Q11424	Film
wd:Q38111	Leonardo DiCaprio

Source: Michael Galkin https://migalkin.github.io/talks/2020-12-04-kgqa

Knowledge Graph-based Question Answering
Pre-defined SPARQL templates

Pros

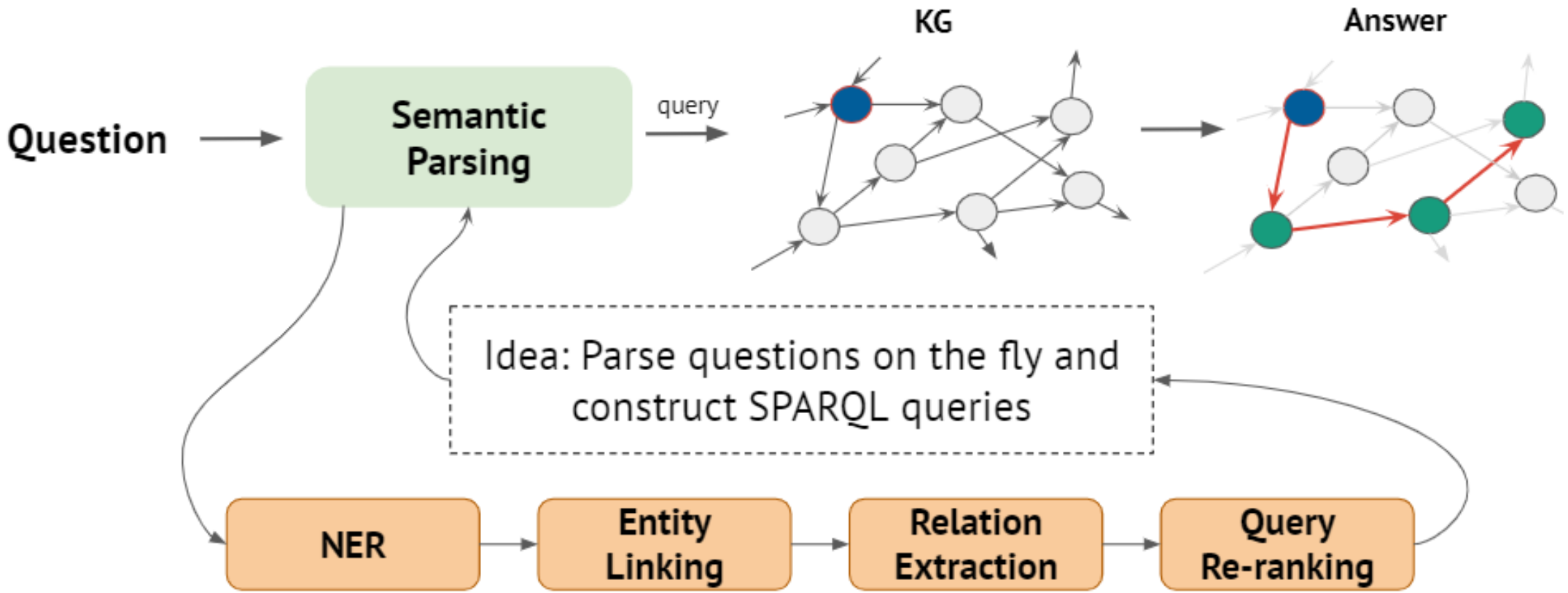
- Independent of the KG size
- Fast & parallelizable
- Explainable query results

Cons

- **Manual** curation of templates (100+ is already hard to sustain)
- Each new question formulation will require a **new** template
- **Hard-coded** to the KG schema (ontology)

Source: Michael Galkin https://migalkin.github.io/talks/2020-12-04-kgqa

Knowledge Graph-based Question Answering
Pipelines: Natural Language 2 SPARQL



Source: Michael Galkin https://migalkin.github.io/talks/2020-12-04-kgqa

Knowledge Graph-based Question Answering
Pipelines: Natural Language 2 SPARQL

Pros

- Some supervised ML can be applied
- Transfer learning works
- Component ↑ -> Performance ↑
- Explainable query results

Cons

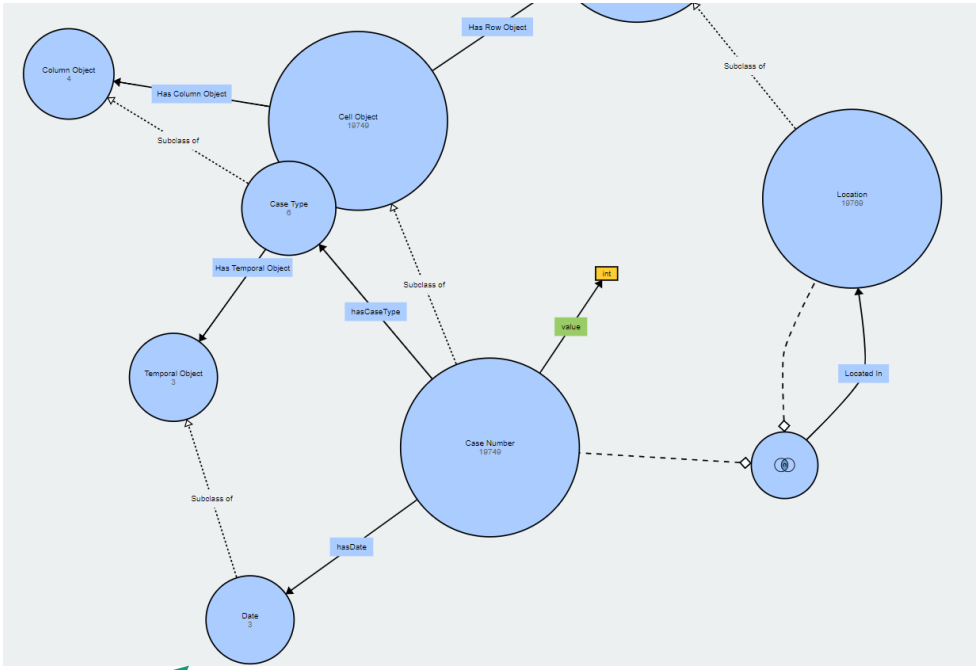
- **Fast** retrieval & communication to the KG is **essential**
- A **Snowball effect** of components error propagation
- **Brute-force** heuristics, e.g., extract a 2-hop subgraph & rank; extract all k-long relation paths and rank

Source: Michael Galkin https://migalkin.github.io/talks/2020-12-04-kgqa

Knowledge Graph-based Question Answering

COVID-19 cases

```
FIPS,Admin2,Province_State,Country_Region,Last_Update,Lat,Long_,Confirmed,Deaths,Recovered,Active,Combined_Key,Incident_Rate,Case_Fatality_Ratio
,,,Afghanistan,2020-12-12 05:26:19,33.93911,67.709953,48116,1945,38141,8030,Afghanistan,123.60146557491366,4.042314406850112
,,,Albania,2020-12-12 05:26:19,41.1533,20.1683,46863,977,24136,21750,Albania,1628.4314406838555,2.0848003755628106
,,,Algeria,2020-12-12 05:26:19,28.0339,1.6596,91121,2575,59590,28956,Algeria,207.7966537762853,2.825912797269565
,,,Andorra,2020-12-12 05:26:19,42.5063,1.5218,7236,78,6598,560,Andorra,9365.171811298776,1.077943615257048
,,,Angola,2020-12-12 05:26:19,-11.2027,17.8739,16061,365,8798,6898,Angola,48.8677327161088,2.2725857667642115
,,,Antigua and Barbuda,2020-12-12 05:26:19,17.0608,-61.7964,147,4,138,5,Antigua and Barbuda,150.11028510742585,2.7210884353741496
,,,Argentina,2020-12-12 05:26:19,-38.4161,-63.6167,1489328,40606,1324792,123930,Argentina,3295.28132683724,2.726464553140745
,,,Armenia,2020-12-12 05:26:19,40.0691,45.0382,146317,2445,123474,20398,Armenia,4937.7470695868105,1.6710293404047376
,,Australian Capital Territory,Australia,2020-12-12 05:26:19,-35.4735,149.0124,117,3,114,0,"Australian Capital Territory, Australia",27.330063069
,,New South Wales,Australia,2020-12-12 05:26:19,-33.8688,151.2093,4639,53,3191,1395,"New South Wales, Australia",57.14461690071447,1.142487605087
,,Northern Territory,Australia,2020-12-12 05:26:19,-12.4634,130.8456,62,0,52,10,"Northern Territory, Australia",25.2442996742671,0.0
,,Queensland,Australia,2020-12-12 05:26:19,-27.4698,153.0251,1226,6,1199,21,"Queensland, Australia",23.966376698269965,0.4893964110929853
,,South Australia,Australia,2020-12-12 05:26:19,-34.9285,138.6007,562,4,558,0,"South Australia, Australia",31.995445488186736,0.7117437722419929
,,Tasmania,Australia,2020-12-12 05:26:19,-42.8821,147.3272,234,13,217,4,"Tasmania, Australia",43.69747899159664,5.555555555555555
,,Victoria,Australia,2020-12-12 05:26:19,-37.8136,144.9631,20350,820,19525,5,"Victoria, Australia",306.94278948400427,4.0294840294840295
,,Western Australia,Australia,2020-12-12 05:26:19,-31.9505,115.8605,835,9,815,11,"Western Australia, Australia",31.74180795255835,1.077844311377
,,,Austria,2020-12-12 05:26:19,47.5162,14.5501,316581,4289,273503,38789,Austria,3515.0670634215667,1.3547875583184081
,,,Azerbaijan,2020-12-12 05:26:19,40.1431,47.5769,167155,1840,103550,61765,Azerbaijan,1648.6055325014115,1.10077473004098
,,,Bahamas,2020-12-12 05:26:19,25.025885,-78.035889,7623,163,6043,1417,Bahamas,1938.471397184474,2.138265774629411
,,,Bahrain,2020-12-12 05:26:19,26.0275,50.55,88820,347,86867,1606,Bahrain,5219.845285243212,0.3906777752758388
,,,Bangladesh,2020-12-12 05:26:19,23.685,90.3563,487849,6986,414318,66545,Bangladesh,296.22370981862264,1.432000475556986
,,,Barbados,2020-12-12 05:26:19,13.1939,-59.5432,292,7,269,16,Barbados,101.61080971983951,2.3972602739726026
,,,Belarus,2020-12-12 05:26:19,53.7098,27.9534,156359,1246,133930,21183,Belarus,1654.7114866771908,0.7968840936562653
,,Antwerp,Belgium,2020-12-12 05:26:19,51.2195,4.4024,71368,0,0,71368,"Antwerp, Belgium",3841.148426306765,0.0
,,Brussels,Belgium,2020-12-12 05:26:19,50.8503,4.3517,78152,0,0,78152,"Brussels, Belgium",6466.635003169107,0.0
,,East Flanders,Belgium,2020-12-12 05:26:19,51.0362,3.7373,57603,0,0,57603,"East Flanders, Belgium",3802.0176045368376,0.0
,,Flemish Brabant,Belgium,2020-12-12 05:26:19,50.9167,4.5833,45941,0,0,45941,"Flemish Brabant, Belgium",4008.201190917617,0.0
,,Hainaut,Belgium,2020-12-12 05:26:19,50.5257,4.0621,96379,0,0,96379,"Hainaut, Belgium",7169.770896736522,0.0
```



COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University

https://github.com/CSSEGISandData/COVID-19

Map

Fraunhofer IAIS

What is your question?

Senden

Examples

How many cases are in total in Germany till 25th August 2020?

How many new cases were found on 24th August 2020 in Argentina?

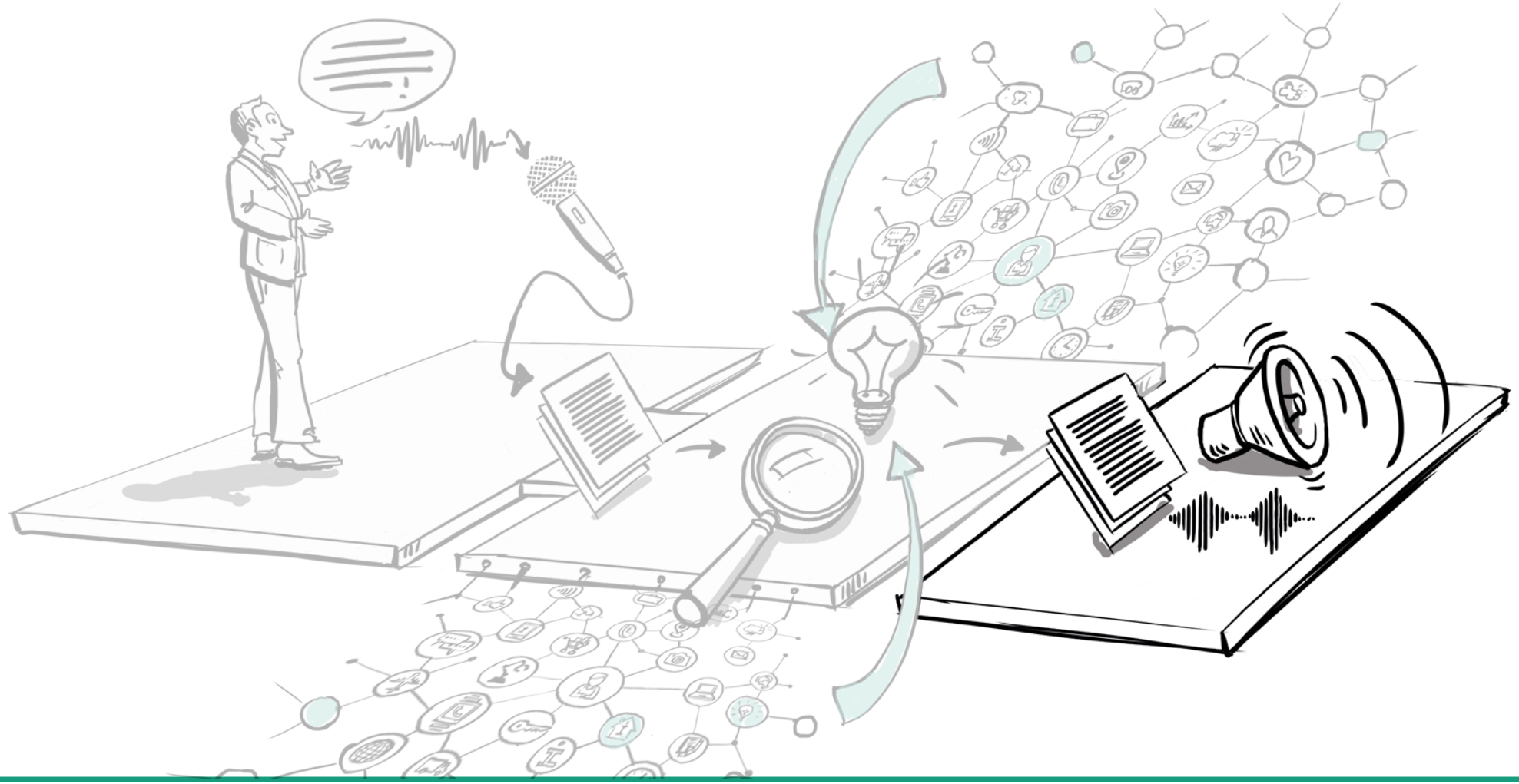
What country has the maximum cases in the world?

Number of recover people in Mexico?

Are there new cases in India?

Which country had the most cases on 24th August 2020?

Speech Synthesis (Text to Speech)



Text to Speech Examples

»Yes, ask me about this city…«

»Technical University Berlin…«

»The Museum is subdivided into the…«

Nordwind und Sonne

»It depends on my job, but I really…«

»En beherztet Kölle Allaaf…«

Conversational AI & Knowledge Graphs
Conclusions

- Increasing acceptance of conversational methods as universal interfaces
- We are creating Conversational AI technology empowered by Knowledge Graphs
- SPEAKER is primarily targeting industry (B2B focus rather than B2C as Siri, Google Assistant, Alexa)
- Our technology pursuit digital sovereignty: Control over own data as well as used technology, Compliance with German and European regulations (e.g. for cloud services), Security and privacy for company data

Thank you!

Danke!

Gracias!

Merci!