First International TEXT2SPARQL Challenge

Co-Located with Text2KG at ESWC25

Dr. Sebastian Tramp

CTO @ eccenca GmbH

Agenda

- 10:40 - **Sebastian Tramp**: Welcome and Introduction to the Challenge
- 10:50 - Jan Wardenga and **Tobias Käfer**: Presentation of the Institute of Applied Informatics and Formal Description Methods (AIFB) @ KIT
- 11:05 - Mehrzad Shahinmoghadam, **Tommaso Soru** and Sanju Tiwari: Presentation of the DBpedia Group
- 11:20 - Daniel Henselmann, **Rene Dorsch** and Andreas Harth: Presentation of the Fraunhofer Institute for Integrated Circuits (IIS)
- 11:35 - Daniel Gerber, Lorenz Bühmann, **Lars-Peter Meyer**, Felix Brei and Claus Stadler: Presentation of the ETi @ Institute for Applied Informatics (InfAI)
- 11:50 - **Oleg Somov**, Daniil Berezin and Roman Avdeev: Presentation of the Artificial Intelligence Research Institute (AIRI), Moscow - remote
- 12:05 - Aleksandr Perevalov and **Andreas Both**: Presentation of the WSE Research Group @ Leipzig University of Applied Sciences
- 12:20 - **Edgard Marx** and Sebastian Tramp: Result Presentation / Feedback Session and Winner Ceremony
- 12:45 - Closing

Overview

- 27 Participants
- 14 Affiliations from 6 Countries
- 12 Endpoints registered
- 7 Presentations
- 3 Withdrawals (after result publication)
- 2 Question Datasets
- 250 Questions

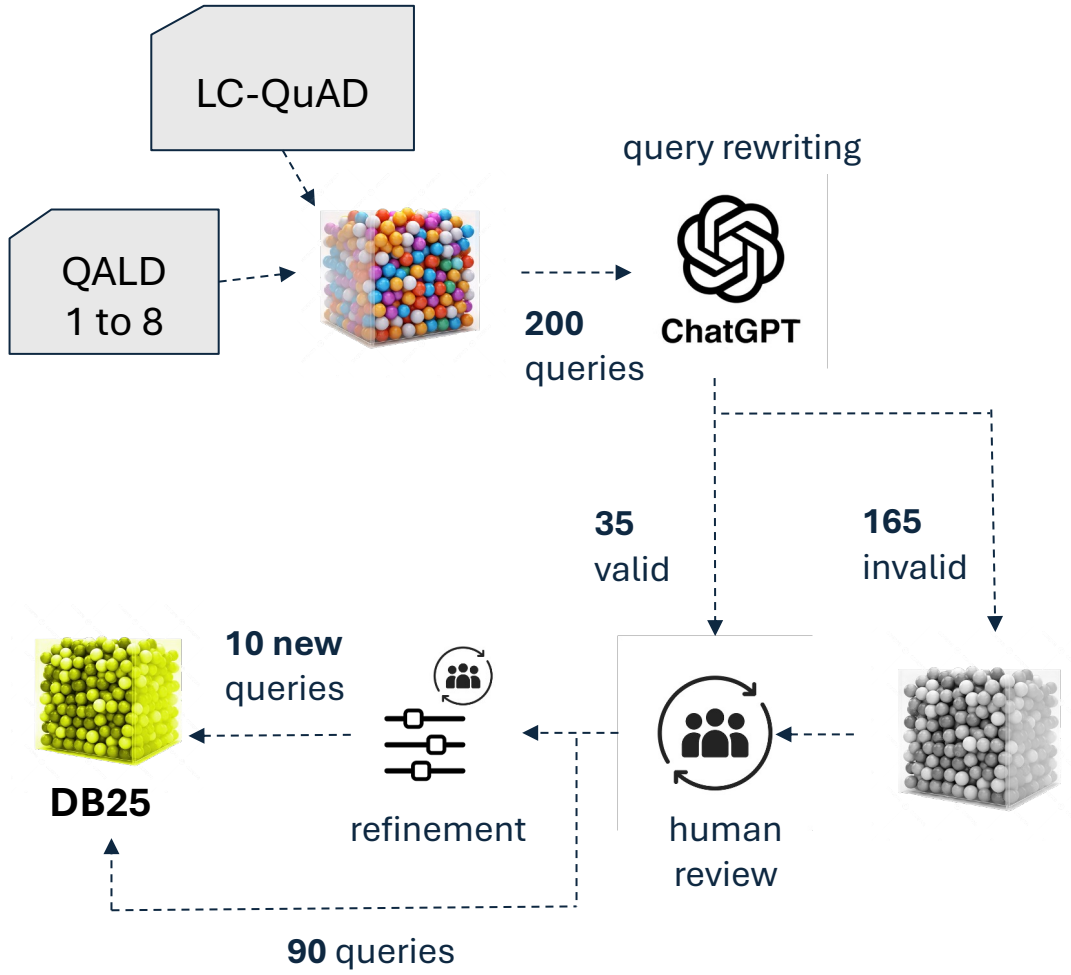
Question Datasets

- DB25 – Dbpedia
  - 200 Questions
    - 100 in english
    - 100 in spanish
- CK25 – Corporate Knowledge
  - 50 Questions in english

```yaml
- id: 15
  question:
    en: What is the cheapest Encoder we can get from a french or german supplier?
  features:
    - SELECT
    - ORDER
    - FILTER
    - LIMIT
  classes:
    - :Product
    - :ProductCategory
    - :Supplier
    - :Price
  properties:
    - :hasCategory
    - :price
    - :amount
    - :hasSupplier
    - :addressCountry
  query:
    sparql: |
      PREFIX pv: <http://ld.company.org/prod-vocab/>
      SELECT DISTINCT ?result
      WHERE
      {
        ?result pv:hasCategory <http://ld.company.org/prod-instances/prod-cat-Encoder> .
        ?result pv:price ?priceR .
        ?priceR pv:amount ?price .
        ?result pv:hasSupplier ?supplier .
        ?supplier pv:addressCountry ?country .
        FILTER ( (?country = "France") ||  (?country = "Germany") )
      }
      ORDER BY ASC(?price)
      LIMIT 1
```

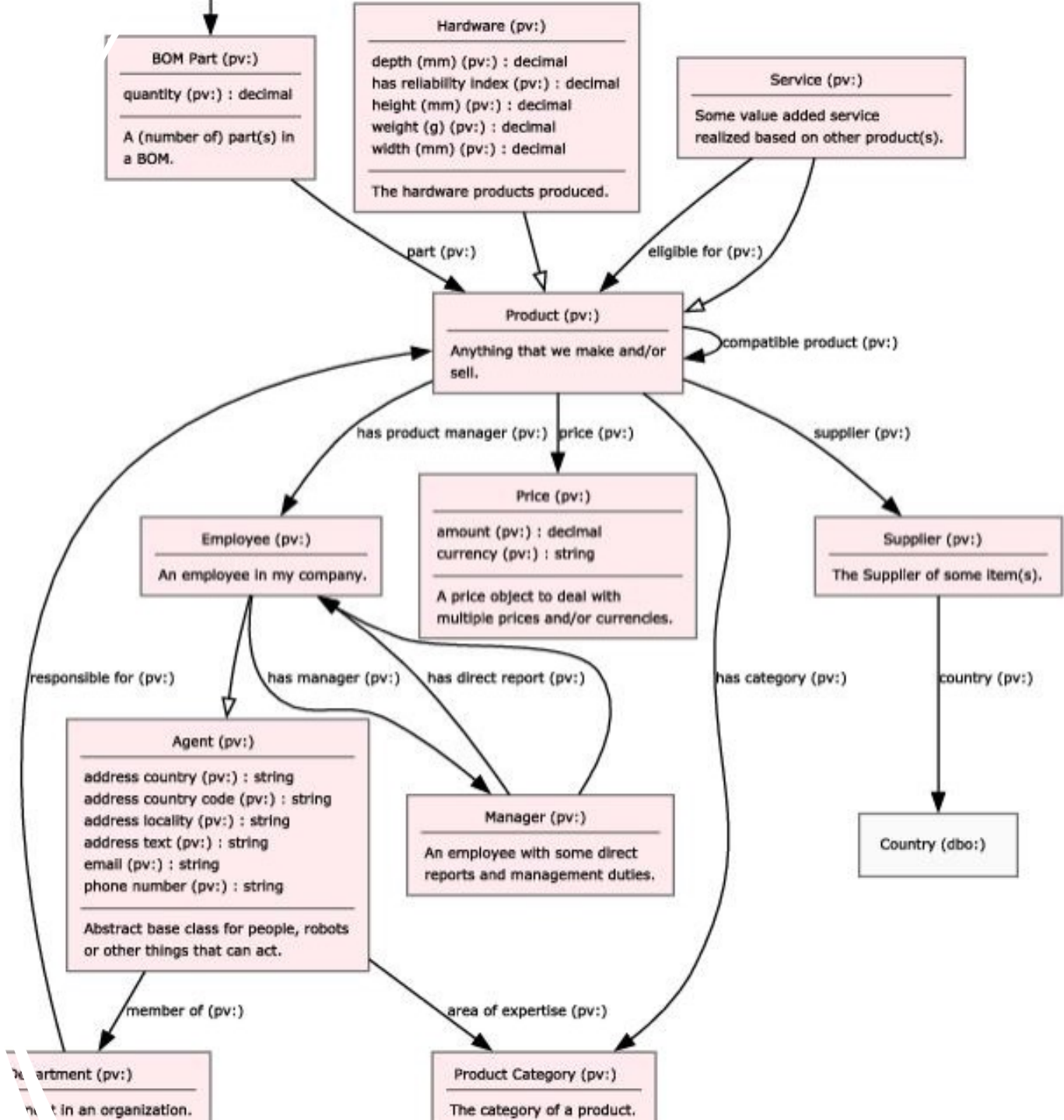
DB25 DBpedia Dataset

- We randomly generated 200 questions and queries using QALD-1 to 8 + LC-QuAD (en + es)
- Out of the 200, 35 returned proper results on the first try.
- The other 65 we had to fix from the remaining 165
  - We selected 100 and reviewed and corrected all of them.
  - We refined 10 questions with GROUP BY and ORDER BY to make the final benchmark balanced in terms of query features.



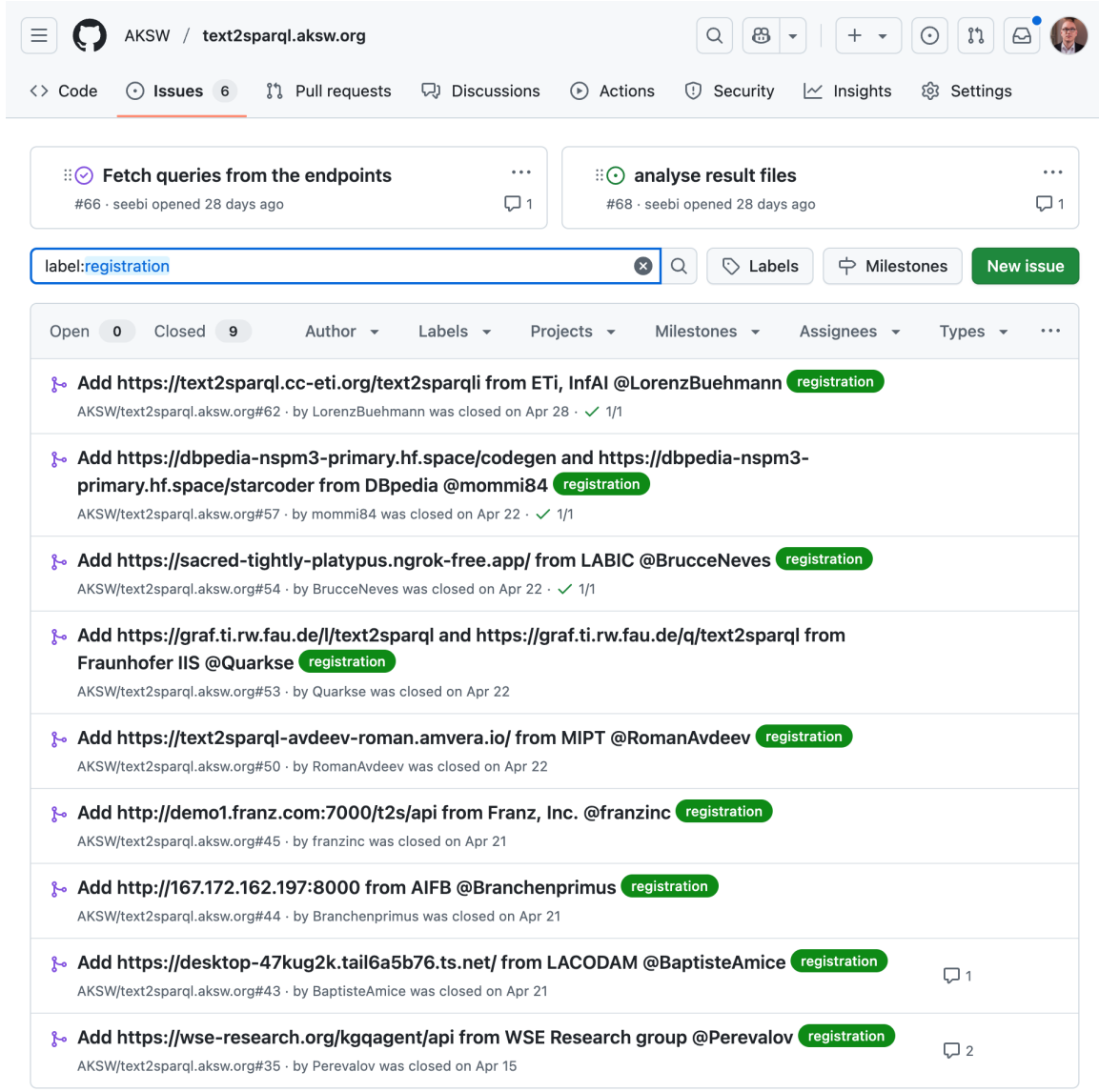
CK25 Corporate Dataset

- Showcase Ontology from the Scratch
- Resource Generation with eccenca Corporate Memory
- Manual Creation of 50 Questions after Interviews with typical Stakeholders



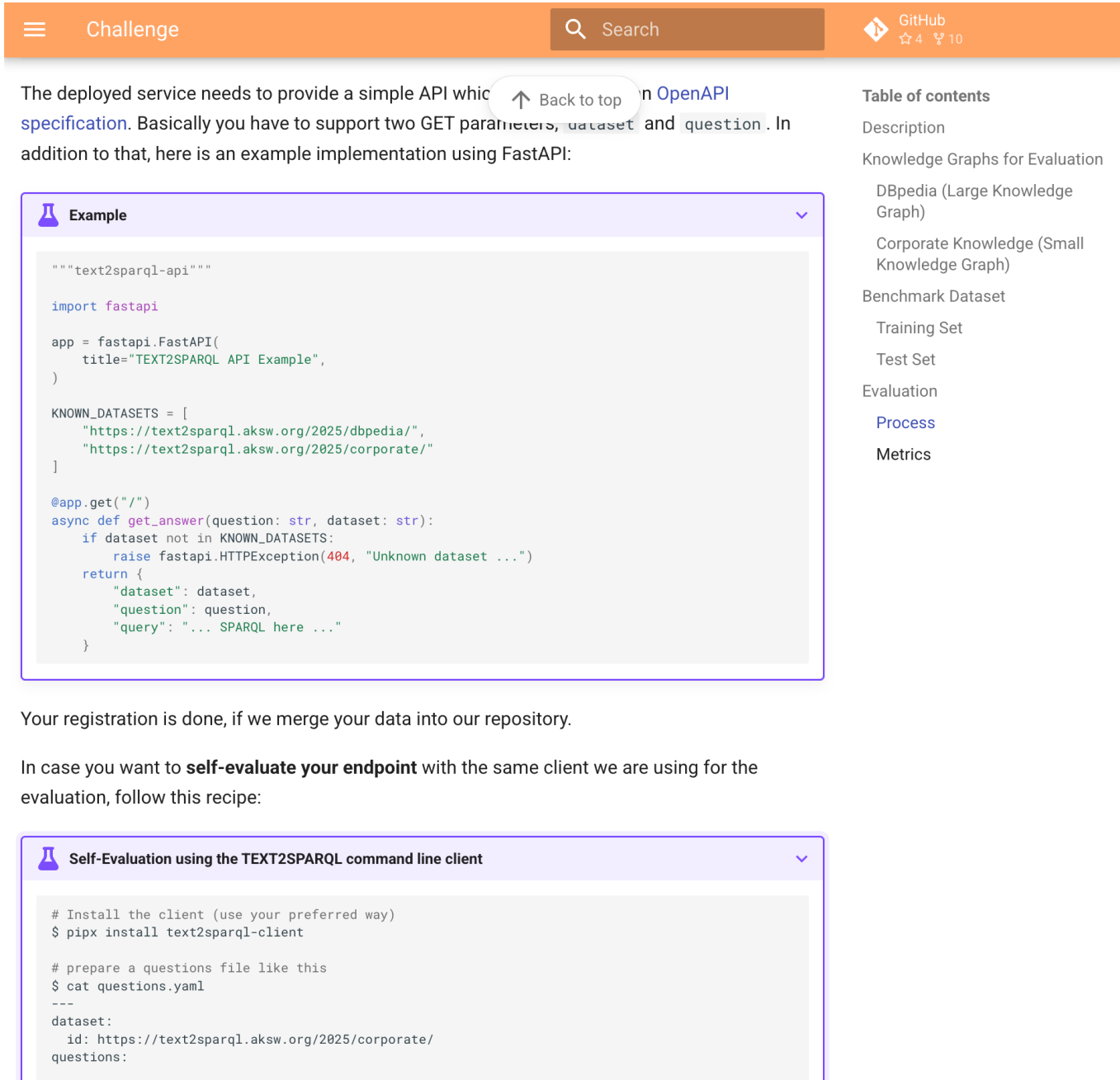
Procedure

- **Challengers had to register upfront**
- We requested to provide an HTTP(S) endpoint with a certain (easy to implement) interface
- We developed a command line client to ask questions on these endpoints
- In the 5 days evaluation phase we sent 250 requests to each endpoint
  - We recorded not only the result content but also the complete response object



Procedure

- Challengers had to register upfront
- **We requested to provide an HTTP(S) endpoint with a certain (easy to implement) interface**
- We developed a command line client to ask questions on these endpoints
- In the 5 days evaluation phase we sent 250 requests to each endpoint
  - We recorded not only the result content but also the complete response object

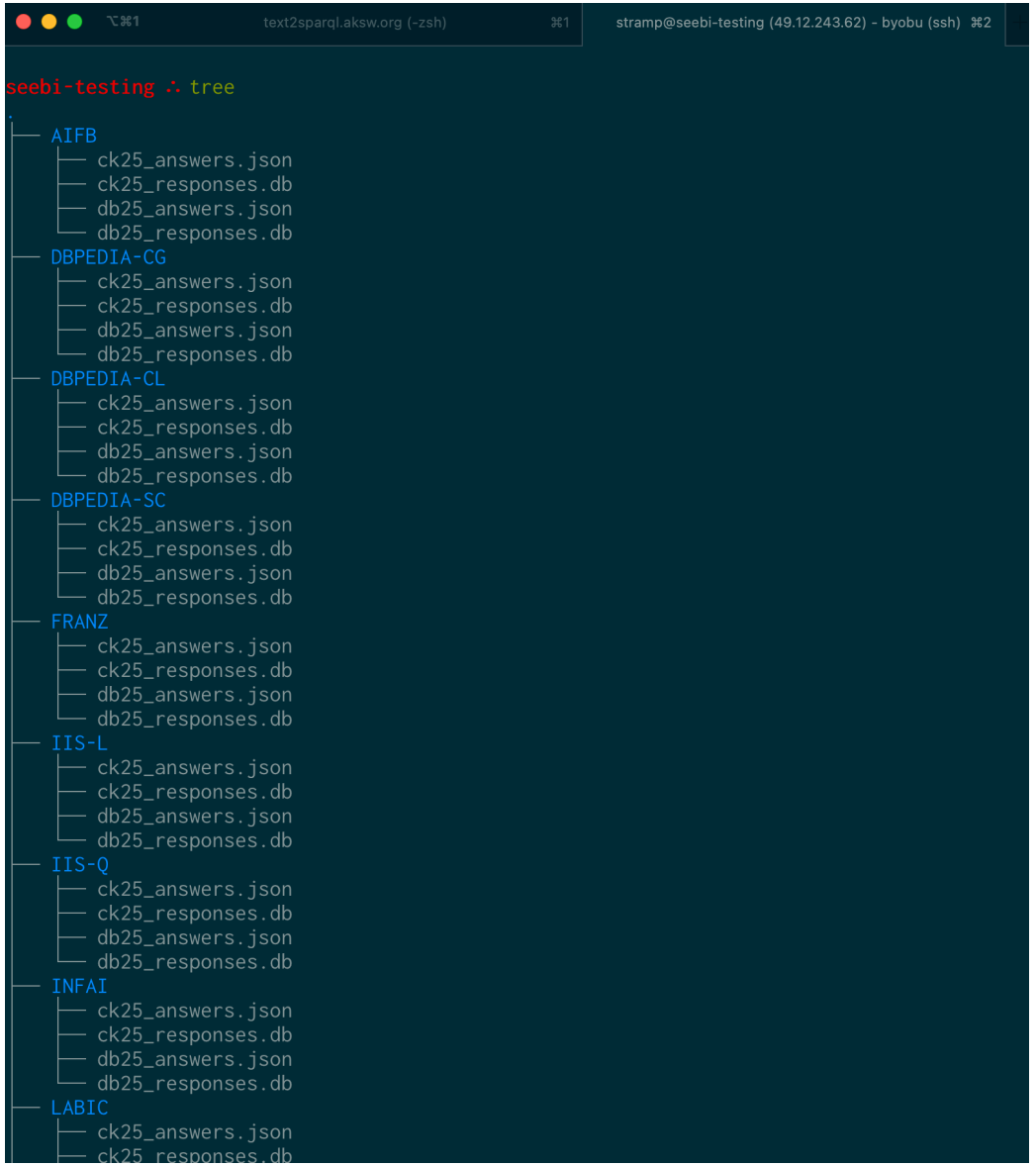


Procedure

- Challengers had to register upfront
- We requested to provide an HTTP(S) endpoint with a certain (easy to implement) interface
- **We developed a command line client to ask questions on these endpoints**
- In the 5 days evaluation phase we sent 250 requests to each endpoint
  - We recorded not only the result content but also the complete response object

Procedure

- Challengers had to register upfront
- We requested to provide an HTTP(S) endpoint with a certain (easy to implement) interface
- We developed a command line client to ask questions on these endpoints
- **In the 5 days evaluation phase we sent 250 requests to each endpoint**
  - We recorded not only the result content but also the complete response object



Agenda

- 10:40 - **Sebastian Tramp**: Welcome and Introduction to the Challenge
- 10:50 - Jan Wardenga and **Tobias Käfer**: Presentation of the Institute of Applied Informatics and Formal Description Methods (AIFB) @ KIT
- 11:05 - Mehrzad Shahinmoghadam, **Tommaso Soru** and Sanju Tiwari: Presentation of the DBpedia Group
- 11:20 - Daniel Henselmann, **Rene Dorsch** and Andreas Harth: Presentation of the Fraunhofer Institute for Integrated Circuits (IIS)
- 11:35 - Daniel Gerber, Lorenz Bühmann, **Lars-Peter Meyer**, Felix Brei and Claus Stadler: Presentation of the ETi @ Institute for Applied Informatics (InfAI)
- 11:50 - **Oleg Somov**, Daniil Berezin and Roman Avdeev: Presentation of the Artificial Intelligence Research Institute (AIRI), Moscow - remote
- 12:05 - Aleksandr Perevalov and **Andreas Both**: Presentation of the WSE Research Group @ Leipzig University of Applied Sciences
- 12:20 - **Edgard Marx** and Sebastian Tramp: Result Presentation / Feedback Session and Winner Ceremony
- 12:45 - Closing