

jobsta

*Study - Job - Money*

Project for  
Introduction to Semantic Systems  
(188.399-2019W)

Group 01

Cem Bicer (01425692)

Helmuth Breitenfellner (08725866)

Laszlo Kiraly (09227679)

Gerald Weber (00125536)

2020-01-31

# jobsta

*Study - Job - Money*

- ▶ For Software Developers and Data Scientists
- ▶ Asks for experience, age, location
- ▶ Answers:
  - ▶ *What shall I study?*
  - ▶ *Where shall I work?*
  - ▶ *What shall I practise?*
  - ▶ *How could I practise?*

- ▶ **Kaggle User Survey**

Data Scientists, Country, Job Role, Programming Language, Income

- ▶ **StackOverflow User Survey**

Software Developer, Country, Job Role, Programming Language, Income

- ▶ **GitHub Repositories Data**

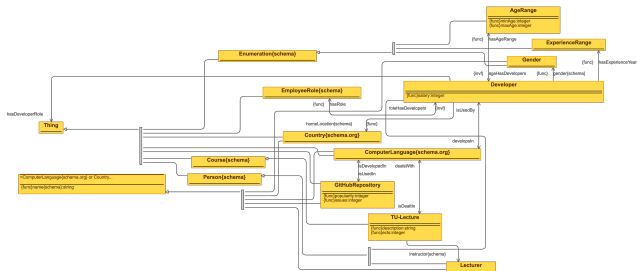
Repository URL, Popularity, Programming Language, Issues

- ▶ **TISS Lectures**

Lectures, Lecturer, Description, Programming Language

Group 01

Cem Bicer  
(01425692)  
Helmuth  
Breitenfellner  
(08725866)  
Laszlo Kiraly  
(09227679)  
Gerald Weber  
(00125536)



- ▶ Created with Protégé
- ▶ Reusing existing Ontologies
  - ▶ <https://schema.org/Gender>
  - ▶ <https://schema.org/ComputerLanguage>
  - ▶ <https://schema.org/EmployeeRole>
  - ▶ <https://schema.org/Country>
  - ▶ ...

# Data Processing #1

jobsta

Group 01

Cem Bicer  
(01425692)  
Helmuth  
Breitenfellner  
(08725866)  
Laszlo Kiraly  
(09227679)  
Gerald Weber  
(00125536)

## Kaggle Survey

- ▶ <https://www.kaggle.com/c/kaggle-survey-2019>
- ▶ Used Jupyter Notebook for Pre-Processing
- ▶ Created RDF XML directly from Python
- ▶ **Challenge:** high number of one-hot-encoded values, had to extract unique values

## StackOverflow Survey

- ▶ <https://insights.stackoverflow.com/survey/2018>
- ▶ Used Jupyter Notebook for Pre-Processing
- ▶ Created RDF XML directly from Python
- ▶ **Challenge:** Fighting with OpenREFINE - at the end reverted to manual

## GitHub Repositories Data

- ▶ <http://ghtorrent.org/>
- ▶ Used Bash & R Script for Pre-Processing
- ▶ Created RDF XML directly from Python
- ▶ **Challenge:** *Huge* data archive (>100GB) had to be filtered / preprocessed

## TISS Lectures

- ▶ <https://tiss.tuwien.ac.at/course/courseList.xhtml?dswid=6403&dsrid=238>
- ▶ Used Python Script
- ▶ Created RDF XML directly from Python (using `rdflib`)
- ▶ **Challenge:** web scraping, identifying the programming language from text

# SPARQL Queries

jobsta

Group 01

Cem Bicer  
(01425692)

Helmuth  
Breitenfellner  
(08725866)

Laszlo Kiraly  
(09227679)

Gerald Weber  
(00125536)

# Questions?

jobsta

Group 01

Cem Bicer  
(01425692)

Helmuth  
Breitenfellner  
(08725866)

Laszlo Kiraly  
(09227679)

Gerald Weber  
(00125536)