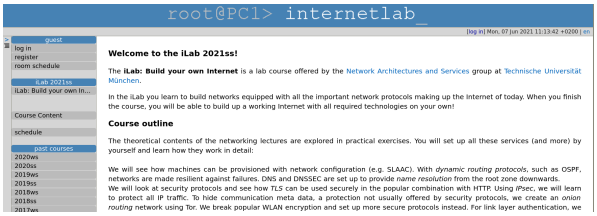
Chair of Network Architectures and Services
Department of Informatics
Technical University of Munich

Thesis B.Sc.

IDP

Similarity Detection for Databases

Motivation



Practical courses conducted at the chair rely on the iLab-Labsystem [1] for the submission of lab answers of course participants. A major learning goal of the courses is to teach participants the ability to explain concepts and results in their own words. Currently, teaching assistants check this through manual work, i.e. check given sources and compare answers within a class. However, comparing answers to answers from previous course iterations is not feasible. The problem is further amplified as the pool of given answers increases considerably with each term.

Topic

The goal of this thesis is to develop and implement a mechanism to compare answers from different classes concerning their similarity. Therefore, methods for plagiarism-checks and methods for similarity checks for free-text answers need to be analyzed. Based on the insights, a scalable mechanism for comparing answers of iLab participants to the answer pool of previous terms should be created. Functionality for the pipeline should:

- automatically identify similar questions between lab instances
- efficiently and scalably fetch and merge data from different databases
- compare large amounts of strings with respect to their similarity

Your Task

- Analysis of plagiarism detection tools
- Analysis of the iLab lab system
- Implementation of an own scalable answer-comparison pipeline running in conjunction to the iLab-lab system

Requirements

- Ability to write efficient and maintainable code
- Participated in any of the iLab courses (preferably iLab1)

Sources

- [1] iLab labsystem, https://github.com/m-o-p/labsystem

Contact

Lars Wüstrich wuestrich@net.in.tum.de
Florian Wiedner wiedner@net.in.tum.de
http://go.tum.de/080755