OSCARS
Open Science Clusters' Action for Research & Society

Funded Project

Opravidlo 2.0

Presenter: Aleš Horák, Faculty of Informatics, Masaryk University, iD 0000-0001-6348-109X

Implemented by

MASARYK UNIVERSITY

Funded by the European Union

What problem(s) are you going to solve?

- Opravidlo is a proofreading service for Czech with explanations
- Currently, it is rule-based with high precision, low recall, and speed issues
- Currently, its UI exists only in Czech, although it is used by language learners

- We will solve the above issues by
  - Using AI models to increase the recall, while keeping high precision
  - Providing explanations and links to Internet Language Reference Book
  - Internationalizing the UI to ease the use by language learners

What are you planning to do to solve the problem?

- AI models for grammar error detection, correction, and explanation
- Possible integration with the rule-based system
- Explanation with reference to the Internet Language Reference Book
- Deployment of the system as a public free service
- Creation of a benchmark of grammar correction from users’ inputs
- A method how to transfer the proofreading service to different languages

What will be the results and how do you plan to make them available to the broader community?

- Social impact: free public proofreading service
    - Already existing, with thousands of users (>200,000 requests per month)
- Scientific impact
    - Service and benchmark data available in large research infrastructures repositories (particularly, LINDAT/CLARIAH-CZ)
    - Dissemination via infrastructures standards (e.g., SSHOC Marketplace, CLARIN Switchboard)

RISKS

What risks could limit the success of the project, and how can they be mitigated

- Deployment risks: computationally intensive
  - Mitigation: university computational infrastructure and/or CLARIN computational infrastructure will be used for deployment
- Data privacy in published data
  - Mitigation: pseudonymization prior to publishing, unrelated fragments will be published (risk of private data disclosure is lowered), not all data will be published

TEAM

OSCARS

Who is doing it?

- Aleš Horák 0000-0001-6
- Emma Bednaříková
- Dana Hlaváčková 0000-0002-9918-09
- Zuzana Nevěřilová 0000-0002-7133-92
- Pavel Rychlý 0000-0001-5
- Patrik Stano
- Tereza Vrabcová 0009-0009-5674-38
- Hana Žižková 0000-0002-6