OSCARS
Open Science Clusters' Action for Research & Society

Funded Project

The ONTOLISST project on the use of NLP for standardized topic assignment in social science repositories

Presenter: Róza Vajda, HUN-REN Centre for Social Sciences, [ORCID: 0000-0002-5325-4536]

Implemented by

HUN REN Centre for Social Sciences

Tampere University

Finnish Social Science Data Archive

SciencesPo Centre de Données Socio-Politiques

poltextLAB

Centre for Social Sciences Research Documentation Centre tkkdk

cnrs

Funded by the European Union

Project goal: *improve visibility of, and access to, social science research*

Problem

Diversity of metadata ontologies used in social sciences impedes data discoverability, accessibility and interoperability.

Solution

ONTOLISST provides AI-assisted tools to simplify topic assignment, thus supporting cross-domain data curation and standardization.

Results & exploitation: *study of methods put in practice*

Research on ontologies

- Collect metadata & CVs from RIs (CLOSER, ESS-Sikt, GESIS, ICPSR…)
- Study classification methods (interviews)
- Dissemination

Development of tools

- Simple, easy-to-use multilingual thesaurus (LiSST)
- Gold standard corpus (for NLP models)
- Demo portal, metadata crosswalk

Why we are doing this?

- Understand data curation practices better & improve them
- Contribute to FAIR social science research data management
- Make social science research more visible for a broader audience
- Strengthen scientific collaboration & synergies of knowledge production

Risks and ways to tackle them

- Running out of time → Extend periods/start earlier – allow overlaps
- Deliverables come late → 2 milestones (LiSST, corpus)
- Will our tools be used? → Start gaining users

TEAM

OSCARS

Our team

Alina DANCIU	Sciences Po, France	https://orcid.org/0000-0002-5126-0078
Jieun JEONG	Sciences Po, France	https://orcid.org/0009-0001-1137-6754
Lucie MARIE	Sciences Po, France	https://orcid.org/0000-0002-7045-4433
Mari KLEEMOLA	Tampere Univ., Finland	https://orcid.org/0000-0001-8855-5075
Judit GÁRDOS	RDC, CSS, Hungary	https://orcid.org/0000-0003-1612-2216
Róza VAJDA	RDC, CSS, Hungary	https://orcid.org/0000-0002-5325-4536
Timea VENCZEL	RDC, CSS, Hungary	https://orcid.org/0009-0008-6174-1722
Éva KOVÁCS	RDC, CSS, Hungary	https://orcid.org/0000-0002-4280-9794
Enikő MEISZTERICS	RDC, CSS, Hungary	https://orcid.org/0000-0002-7078-8820
Júlia EGYED-GERGELY	RDC, CSS, Hungary	https://orcid.org/0000-0003-2259-3823
Anna HORVÁTH	RDC, CSS, Hungary	https://orcid.org/0000-0003-3061-9982
Miklós SEBŐK	PolTextLab, CSS, Hungary	https://orcid.org/0000-0003-0595-2951
Barbara BABOLCSAY	PolTextLab, CSS, Hungary	https://orcid.org/0009-0001-1748-1352
Anna TAKÁCS	PolTextLab, CSS, Hungary	https://orcid.org/0009-0005-6301-4615

SciencesPo CENTRE DE DONNÉES SOCIO-POLITIQUES

cnrs

Tampere University

FINNISH SOCIAL SCIENCE DATA ARCHIVE

tkkdk

poltextLAB

THANK YOU!