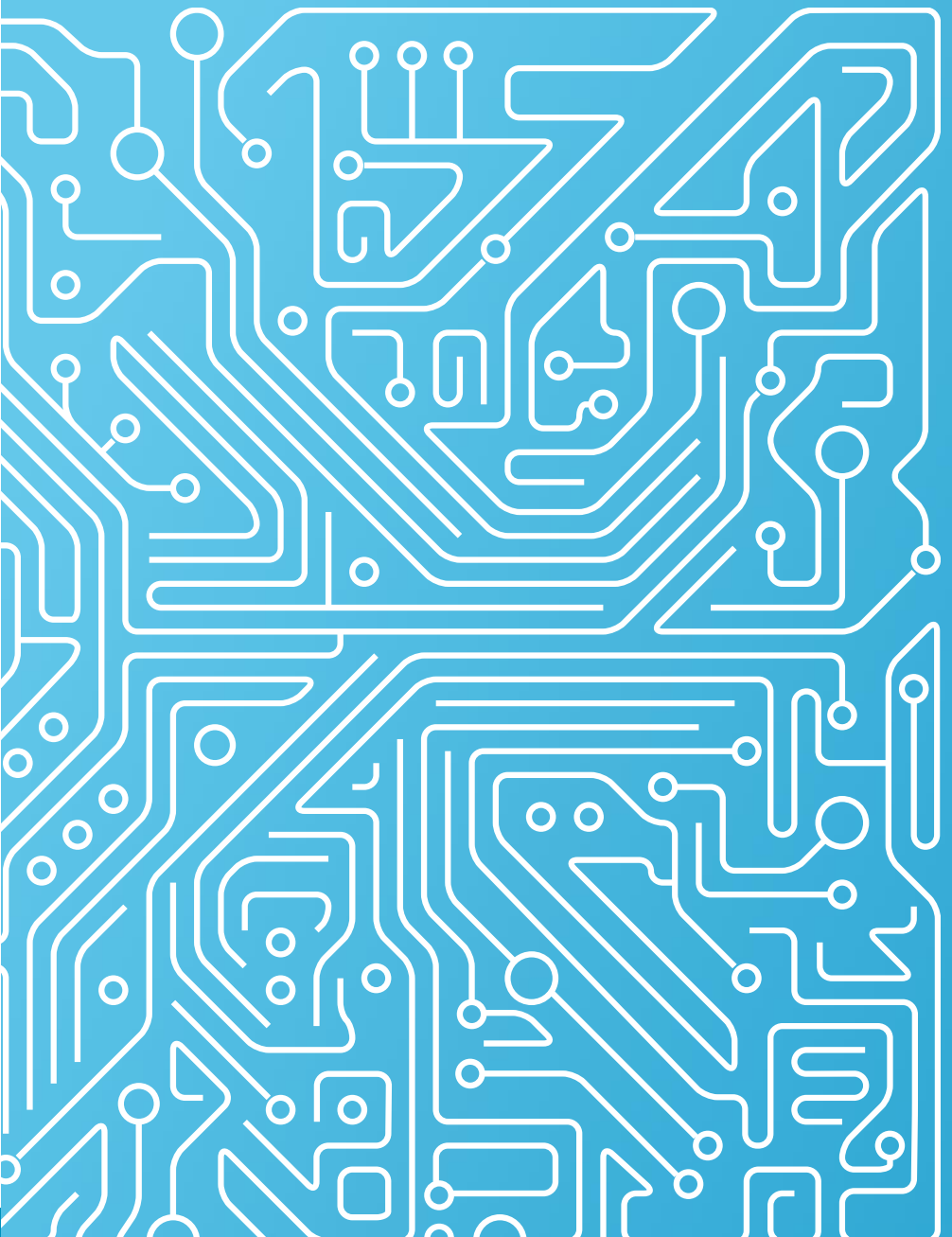


INESC TEC

O PAPEL ESTRATÉGICO DO DATA STEWARD NO DESENVOLVIMENTO DE DMPS EM PROJETOS DE GRANDE DIMENSÃO: O CASO DO INESC TEC

JOÃO AGUIAR CASTRO

FÓRUM GDI – 2024/11/21

Este trabalho é financiado por fundos nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto UIDB/50014/2020
DOI 10.54499/UIDB/50014/2020 | https://doi.org /10.54499/uidb/50014/2020

EQUIPA E SERVIÇOS – APOIO À GESTÃO

Isabel Macedo
Responsável AG

João Aguiar Castro
Data Steward

Inês Sousa
Gestão de Informação

Sofia Ribeiro
Gestão de Informação

Apoio à Decisão
Gestão de Informação
Melhoria Contínua
Gestão de Dados de Investigação

+890
INVESTIGADORES INTEGRADOS

+380
DOUTORADOS

DESENVOLVIMENTO DE DMP

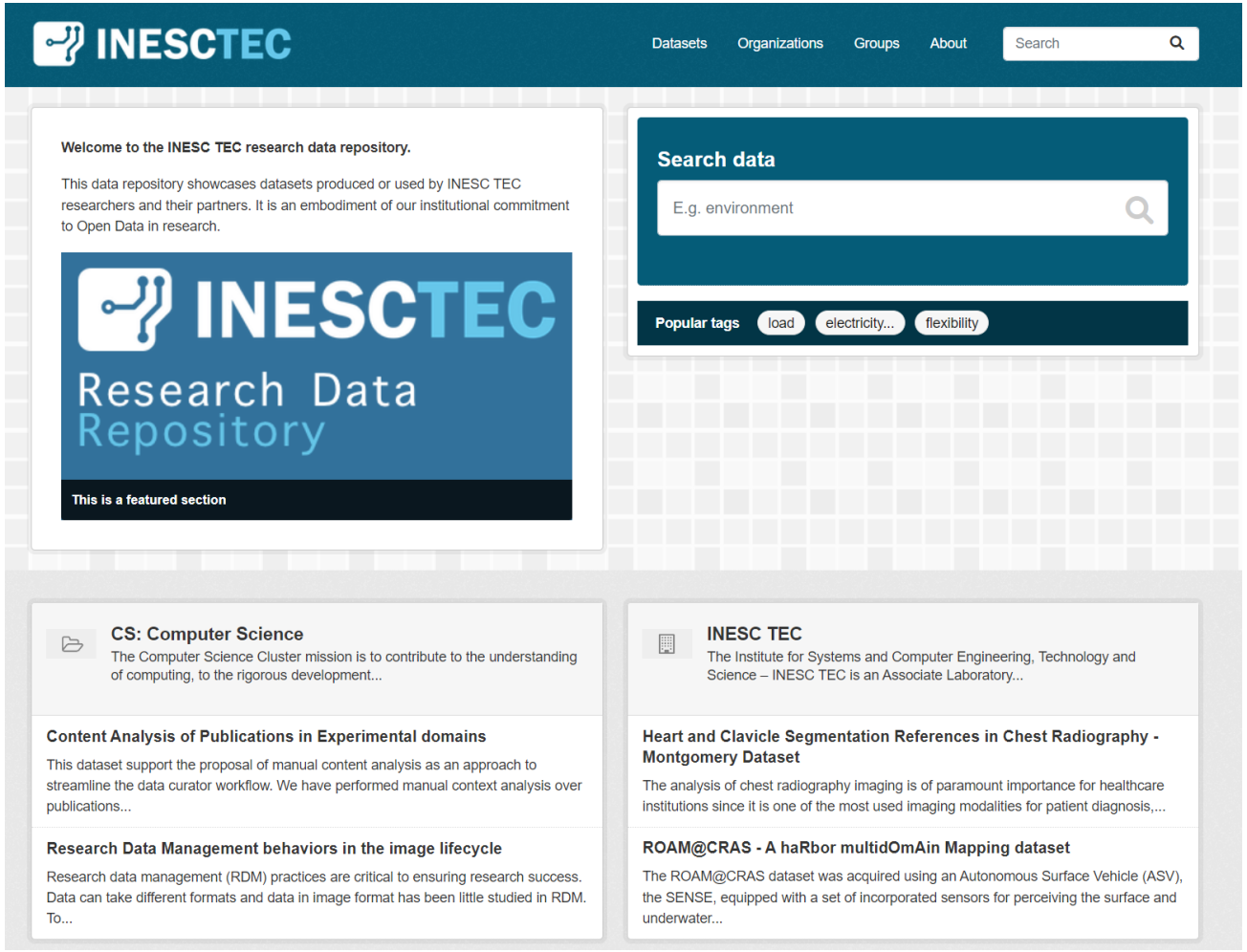
Abordagem no INESC TEC

- Mapeamento de boas práticas e alinhamento com as componentes estruturantes do DMP.
- Apresentação dos objetivos na kick-off meeting dos projetos.
- Template para recolha de informação detalhada sobre os dados que vão suportar o projeto.
- Documento vivo (mas também de trabalho).

Componentes DMP

- **Sumário dos dados**
- **Dados FAIR**
- **Escolha de repositório**
- **Documentação dos dados e metadados**
- **Convenção para nomear ficheiros**
- **Segurança dos dados**
- **Gestão e alocação de recursos**
- **Ethics and Legal Compliance**

APOIO À IMPLEMENTAÇÃO



Classification of online health messages

Classification of online health messages

The dataset has 487 annotated messages taken from Medhelp, an online health forum with several health communities (https://www.medhelp.org/). It was built in a master thesis entitled "Automatic categorization of health-related messages in online health communities" of the Master in Informatics and Computing Engineering of the Faculty of Engineering of the University of Porto. It expands a dataset created in a previous work [see Relation metadata] whose objective was to propose a classification scheme to analyze messages exchanged in online health forums.

A website was built to allow the classification of additional messages collected from Medhelp. After using a Python script to scrape the five most recent discussions from popular forums (https://www.medhelp.org/forums/list), we sampled 285 messages from them to annotate. Each message was classified three times by anonymous people in 11 categories from April 2022 until the end of May 2022. For each message, the rater picked the categories associated with the message and its emotional polarity (positive, neutral, and negative).

Our dataset is organized in two CSV files, one containing information regarding the 885 (=3*285) classifications collected via crowdsourcing (CrowdsourcingClassification.csv) and the other containing the 487 messages with their final and consensual classifications (FinalClassification.csv).

The readMe file provides detailed information about the two .csv files.

Data and Resources

readMe		Explore
Crowdsourcing Classification	Information regarding the 885 (=3*285) classifications collected via...	Explore
Final Classification	487 messages with their final and consensual classifications.	Explore

Additional Info

Field	Value
Source	Medhelp
Author	João Abelha
State	active
Last Updated	July 6, 2022, 2:03 PM (UTC+01:00)
Created	July 6, 2022, 1:55 PM (UTC+01:00)
Citation	Abelha, J. (2022). Classification of online health messages [Data set]. INESC TEC. https://doi.org/10.25747/DG9G-A217
DOI	https://doi.org/10.25747/DG9G-A217
dc.Coverage.Temporal	April - May 2022
dc.File.Size	476 kb
dc.Format	.csv
dc.Relation	Carla Teixeira Lopes and Bárbara Guimarães Da Silva. "A classification scheme for analyses of messages exchanged in online health forums." Proceedings of the The Information Behaviour Conference (ISIC 2018). 2018.

APOIO À IMPLEMENTAÇÃO

zenodo

CONVERGE

The main objective of the CONVERGE project is the development of an innovative toolset aligned with the…

Project https://converge-project.eu/

Institute for Systems and Computer Engineering of Porto, University of Oulu, Barcelona Supercomputing Center, EURECOM, Sorbonne University, INRIA, CSC - IT Center for Science (Finland), allbesmart, Greenerwave (France), adaptech, FINWE, FinCloud.tv, Rice University, Rutgers, The State University of New Jersey, Queen´s University Belfast, InterDigital (United States)

VR2CARE

VR2Care aims to create age-friendly virtual environments fostering the use of interactive technologies for th…

Project https://www.vr2care.eu/

Institute for Systems and Computer Engineering of Porto, AFEDEMY, ALTICE LABS, UNIVERSITA DEGLI STUDI DI NAPOLI FEDERICO II, COGVIS SOFTWARE UND CONSULTING GMBH, COOPERATIVA SOCIALE COOSS MARCHE ONLUS SOCIETA COOPERATIVA PER AZIONI, IMAGINARY SRL, STICHTING SMART HOMES, STICHTING TANTELOUISE, VENERÁVEL ORDEM TERCEIRA DE SÃO FRANCISCO

AI4REALNET

AI4REALNET covers the perspective of AI-based solutions addressing critical systems (electricity, railway,…

Project

Institute for Technological Research SystemX, Fraunhofer Institute for Energy Economics and Energy System Technology, University of Kassel, Polytechnic University of Milan, University of Amsterdam, Delft University of Technology, ZHAW Zurich University of Applied Sciences, University of Applied Sciences and Arts Northwestern Switzerland, Linkoping University, EnliteAI GmbH, Reseau de Transport d´Electricite, TenneT TSO B.V., DB Netz AG, Swiss Federal Railways, NAV Portugal, Institute for Systems and Computer Engineering of Porto

October 17, 2024 (1) | Dataset | Open

Structured Power Grid Simulation Dataset for Machine Learning: Failure and Survival Events in Grid2Op's L2RPN WCCI 2022 Environment

Lehna, Malte; Hassouna, Mohamed; Degtyar, Dmitry; and 2 others

This dataset was developed for and used in the paper titled "Fault Detection for Agents in Power Grid Topology Optimization: A Comprehensive Analysis" by Malte Lehna, Mohamed Hassouna, Dmitry Degtyar, Sven Tomforde, and Christoph Scholz, presented at the Workshop on Machine Learning for Sustainable Power Systems (ML4SPS), part of ECML PKDD 2024. While the pape…

Part of AI4REALNET

Uploaded on October 18, 2024

36 41

March 11, 2024 (v1) | Other | Open

AI4REALNET: Data Management Plan structure and objectives

Aguiar Castro, João

This presentations was developed as part of the AI4REALNET project's Kick off Meeting. The objetive was to raise the awareness of project partners about the importance and objectives of the Data Management Plan.

Part of AI4REALNET, INESC TEC

Uploaded on March 11, 2024

25 26

March 11, 2024 (v1) | Other | Open

AI4REALNET Project presentation: Ethics and Data Protection Committee (EDPC) meeting.

Bessa, Ricardo

This presentation was made as part of the first meeting of the AI4REALNET project's Ethics and Data Protection Committee (EDPC). The Data Protection Officer (DPO) and Ethics Advisors of each partner analysed a draft version of the DMP and provided their recommendations for improvement of the document, and also general recommendations for the pro…

Part of AI4REALNET

Uploaded on March 11, 2024

56 17

APOIO À IMPLEMENTAÇÃO

SpecRF-Posture Dataset

Oliveira, Mariana[1]; Ribeiro, Francisco[1]; Paulino, Nuno[1, 2]; Yurduseven, Okan[3]; Pessoa, Luís[1, 2]

Show affiliations

In recent years, the utilization of Radio Frequency (RF) signals for Human Posture Recognition (HPR) has emerged as a promising approach in wireless sensing technology. Recent studies within RF systems have demonstrated the effectiveness of S21 parameters for human body-related classification tasks. Inspired by these advancements and the enhanced network performance offered by WiFi-6E, our system leverages S21 parameters within this frequency range for posture recognition. By analyzing the S21 parameters, we introduce SpecRF-Posture, a novel system designed for the accurate classification of human posture. Unlike previous approaches, we explore the use of low-cost hardware, based only on passive specular reflections that occur in the path between a steerable horn transmitter, a reflective surface, the Space-of-Interest (SoI), and an omnidirectional receiver. To characterize the region where the individual is positioned using specular reflections, our system achieves beam scanning by mechanically rotating the transmitter at regular intervals. Our work evaluates the viability of utilizing passive scatters within the propagation medium for HPR or similar tasks, at a low hardware cost.

For each posture (Standing, T-shape, Side, Sitting, Lying down), we collected 120 samples (referred t resulting in a dataset of 600 samples (1 subject × 5 postures × 120 sweeps). Each sample is characte (n_resolution, n_angles), where n_resolution=151 represents the number of frequencies between 5.92 resolution of the frequency range, and n_angles = 25 denotes the number of angles through which the of postures included minor movements of the hands or head to introduce variability into the data and r ways of assuming these postures.

Each sweep corresponds to a .csv file where each row corresponds to a frequency and each pair of c and imaginary parts of the S21 parameter measured at each angle of rotation of the transmitter antenn

Additional details

Funding

CONVERGE – Telecommunications and Computer Vision Convergence Tools for Research Infrastructures 101094831

European Commission

Software

Repository URL

https://github.com/franciscombr/SpecRF-Posture

Development Status

Active

UM CASO DE MELHORIA DE DMP COM APOIO DO DATA STEWARD

PEDIDO DE REVISÃO DE DMP POR UM OFICIAL EUROPEU

O QUE PRECISAVA DE SER REVISTO…no espaço de uma semana

1. As funções definidas no âmbito do projeto para gerir os dados devem ser clarificadas.
2. Os aspectos documentais do projeto não estão suficientemente descritos.
3. Os metadados são importantes para a utilização de dados FAIR, mesmo que os dados sejam sensíveis e não possam ser disponibilizados.

"While not to share any sensitive data at all is a viable strategy, it is not the most useful strategy from a research perspective"

O QUE FOI FEITO PARA MELHORAR O DMP

COMPONENTE	PRIMEIRA VERSÃO	VERSÃO DO DATA STEWARD
1. General Information	1.1. Project abstract	1.1. Project abstract **1.2. DMP objectives and scope**
2. Data summary	2.1. Use cases 2.2. Trust Outputs	2.1. Use cases **2.2. Data Inventory Register** 2.3. Trust Outputs
3. FAIR Data	**3.1. Use cases** 3.2. Repositories	3.1. Repository selection for data availability **3.2. Licencing** **3.3. Data identifier** **3.4. Data citation** **3.5. Data documentation** **3.6. File naming convention** **3.7. Metadata standards**
4. Allocation of resources	4.1. Costs for making the data FAIR 4.2. Responsabilities for data management in the project 4.3. Potential value of long-term preservation	
5. Data security		**5. Data security and Backup**
6. Ethical aspects		**6.1. Data Minimization** **6.2. Data Anonymization**

SOBRE O PAPEL ESTRATÉGICO DO DATA STEWARD

Conhecimento transversal das boas práticas GDI agiliza o processo e facilita a vida dos investigadores

Pode desempenhar uma diversidade de papéis em projetos de grande dimensão:

- Desenvolver e assumir o papel de Gestor do DMP
- Apoio contínuo ao investigador
- Representação institucional em atividades formais relacionadas com o DMP
- Revisão de DMPs liderados por outros institutos ou parceiros
- Capacitação dos investigadores

Resposta rápida a pedidos urgentes

– Temos que entregar um Plano de Gestão de Dados esta semana, não sabemos o que é.

Podes ajudar?