Linguistic Linked Open Data for Humanists

Anas Fahad Khan, Giulia Pedonese, Michele Mallia

Lisbon Summer School in Linguistics, July 1-5, 2024

Introduction

Who we are and what we do

The CNR Institute of Computational Linguistics (ILC)

We are researchers and research technologists at the Italian National Research Council (CNR) and we work at the Institute of Computational Linguistics in Pisa

Istituto di Linguistica Computazionale
"Antonio Zampolli"

Consiglio Nazionale delle Ricerche

H2IOSC project

H2IOSC stands for Humanities and Cultural Heritage Italian Open Science Cloud

It is a PNRR project aiming to create a collaborative and inclusive cluster of the Italian nodes of four Research Infrastructures in the domain of Social Sciences and Humanities: CLARIN-IT, DARIAH-IT, E-RIHS.it, and OPERAS-IT

It will provide researchers, businesses and citizens with an open and multidisciplinary environment enabling access to advanced tools to conduct innovative and intensive data-driven research

https://www.h2iosc.cnr.it/

The Institute of Computational Linguistics in H2IOSC

The project comprises various work packages, involving multiple units from different research centers. The ILC is the leading member of CLARIN-IT and hosts the ILC4CLARIN center. Therefore, our role in H2IOSC is to improve and support the use of CLARIN tools and services

Overview of Research Infrastructures in H2IOSC

H2IOSC Work packages

WP1 Project and Financial Management

WP2 Landscaping

WP3 Digital Resources Standardization, Consolidation and Alignment

WP4 Ris Nodes and resources Interoperability

WP5 MarketPlace

WP6 Resources Accessibility, Servification, Virtualization, Remotization

WP7 Community Pilots

WP8 Training and Capacity Building

H2IOSC Training

The units working on WP8 have created a joint strategy to organise training activities for the Italian user community. The goal is to provide them with interdisciplinary skills and knowledge specific to SSH, as well as to train them on the resources available from IR. The strategy involves:

- The adaptation of pre-existing training materials to the needs of the user base

- The creation of a training infrastructure composed of two platforms: a **Virtual Training Environment** (LMS) for delivery and the **FAIR and Interoperable Training Materials Publication Platform** for repository and versioning

Virtual Training Environment Platform

Results of our pre-course survey

1. Are you familiar with the FAIR principles?

Altri dettagli   💡 Dati analitici

🔵	Yes	3
🟠	No	8
🟢	I've heard about them	5

3. Have you already worked with language resources?

🔵	Yes	13
🟠	No	1
🟢	I am not sure what language res...	2
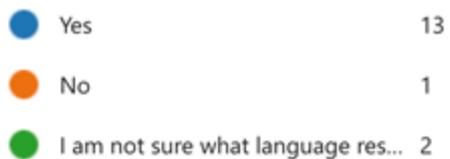
4. If yes, what type of language resources have you already worked with?

🔵	Textual resources (corpora)	12
🟠	Lexical resources (lexica, Diction...	11
🟢	Natural Language Processing to...	3
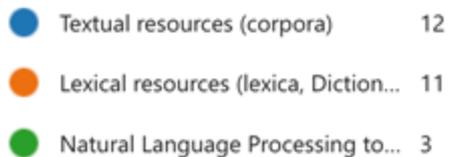
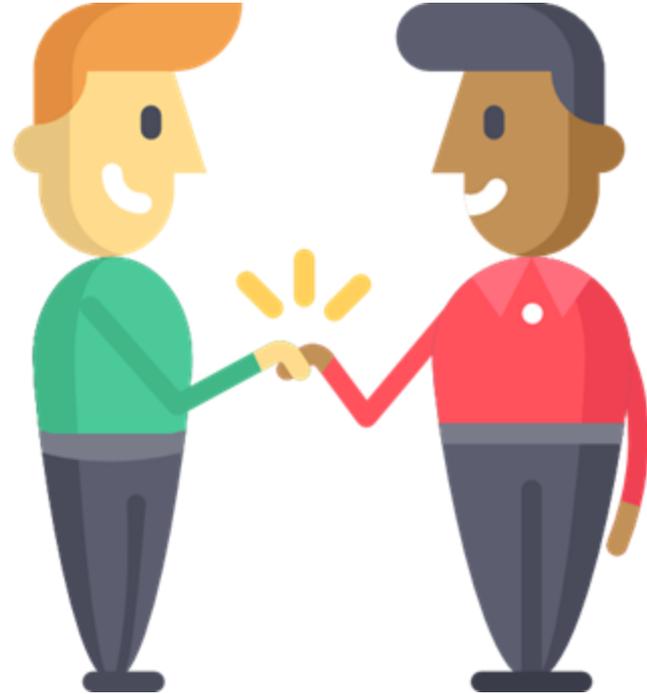9. Which of the following semantic web standards are you familiar with? (Select all that apply)

- 🔵 RDF (Resource Description Fram...   4
- 🟠 OWL (Web Ontology Language)   2
- 🟢 SPARQL (SPARQL Protocol and ...   1
- 🔴 SKOS (Simple Knowledge Organ...   4
- 🟣 None of the above   10

Introductions

- anasfahad.khan@cnr.it
- giulia.pedonese@cnr.it
- michele.mallia@cnr.it

Introduction icons created by Freepik - Flaticon

Your turn!

Introduce yourself in a few words

Course outline

Day 1 – Introduction to LOD: the fundamentals

Day 2 – SPARQL tutorial; Language resources, FAIR principles and LOD

Day 3 – Semantic Web standards and Semantic artefacts: taxonomies, thesauri and ontologies

Day 4 – Tour of the LLOD Cloud and Ontolex; Lexicographic queries using SPARQL

Day 5 – Introducing triplestores and Protégé; LOD case study: the LiLa-Linking Latin project

Let's begin!

We are happy to have you onboard