

Estrazione di dati e storytelling per l'ecosistema dei progetti di Digital Humanities sul patrimonio culturale italiano

codice proposta 20227M8RS7 CUP J53D23013000006

PROGETTO DI RICERCA

L'assegno prevede attività di knowledge extraction, data visualisation e storytelling inerenti il progetto "The ATLAS of Italian Digital Humanities: a dynamic knowledge graph of digital scholarly research on Italian Cultural Heritage" (PRIN 2022). Il progetto intende realizzare un ambiente online per (1) l'estrazione e l'annotazione dei dati della ricerca, (2) la creazione di un knowledge graph e (3) l'esplorazione del patrimonio culturale italiano prodotto ed elaborato in progetti di ricerca internazionale di Digital Humanities.

(1) L'ambiente da realizzare (ATLAS) è basato sull'estensione di software esistenti (CLEF) e prevede interfacce per l'inserimento dati da parte dei ricercatori e contributori anonimi, nonché meccanismi per l'estrazione semi-automatica di ulteriori dati della ricerca segnalati dai contributori. (2) I dati inseriti manualmente contribuiranno alla realizzazione di un grafo della conoscenza dedicato ai progetti internazionali di Digital Humanities devoti al patrimonio culturale italiano. (3) L'ambiente online sarà dedicato ad esperti di dominio (filologi, storici, bibliotecari, archivisti), studenti e curiosi.

Le collezioni digitali e i dati da valorizzare provengono da fonti online (open data) e/o vengono segnalati dai membri del progetto e raccolti dall'assegnista. Tutti i dati devono essere disseminati sotto forma di Linked Open Data (LOD), serviti attraverso SPARQL endpoint o Web API. L'ambiente online che servirà tali dati dovrà essere dotato di meccanismi per l'interrogazione dei dati e la restituzione di interfacce per browsing, esplorazione e la ricerca. Particolare attenzione verrà dedicata alla realizzazione di interfacce per data storytelling, ovvero un ambiente in cui l'utente può interrogare una o più fonti di dati (LOD e non) tramite interfacce user-friendly e ottenere grafici interattivi per la scoperta di informazioni.

Lo sviluppo dell'applicazione finale prevede l'utilizzo e quindi la conoscenza avanzata di linguaggi di programmazione (Python), framework per il web (Flask, webpy), linguaggi per lo sviluppo web (HTML, CSS, Javascript e librerie come JQuery), protocolli e tecnologie lo scambio di dati sul web (HTTP, REST, API) e tecnologie del Semantic Web (RDF, SPARQL, OWL, SPARQLAnything).

PIANO DELLE ATTIVITA'

L'assegnista collaborerà con i membri del progetto per la raccolta dei requisiti funzionali delle interfacce, l'implementazione dei template grafici, l'acquisizione dei dati e la realizzazione di data stories. Per ottemperare agli obiettivi prefissati, le attività previste sono le seguenti:

- Analisi dello stato dell'arte di software per l'estrazione automatica e lo storytelling basati su Linked Open Data
- Sviluppo di un applicativo software per l'annotazione, l'interrogazione, l'estrazione e la visualizzazione dei dati
- Design e implementazione di almeno 3 template grafici (e.g. mappe, timeline, statistiche) e 6 data stories
- Meccanismi di trasformazione, export e salvataggio dei dati
- Messa in produzione sui server del progetto
- Documentazione e disseminazione dei risultati

Knowledge extraction and storytelling from Digital Humanities data on the Italian Cultural Heritage

codice proposta 20227M8RS7 CUP J53D23013000006

BRIEF DESCRIPTION

Candidates are requested to study and perform activities relevant to tasks like knowledge extraction, data visualisation and storytelling in the context of the project “The ATLAS of Italian Digital Humanities: a dynamic knowledge graph of digital scholarly research on Italian Cultural Heritage” (PRIN 2022). The goal of the project is to realise an online environment that allows anyone to (1) extract information from scholarly data, (2) elaborate crowdsourced data and populate a knowledge graph, and (3) explore the Italian Cultural Heritage via dedicated interfaces. The development of the application requires expertise in programming languages, languages for web development, protocols for data exchange on the Web, and Semantic Web technologies.

BREVE DESCRIZIONE

L'assegno prevede attività di knowledge extraction, data visualisation e storytelling inerenti il progetto “The ATLAS of Italian Digital Humanities: a dynamic knowledge graph of digital scholarly research on Italian Cultural Heritage” (PRIN 2022). Il progetto intende realizzare un ambiente online per (1) l'estrazione e l'annotazione dei dati della ricerca, (2) la creazione di un knowledge graph e (3) l'esplorazione del patrimonio culturale italiano prodotto ed elaborato in progetti di ricerca internazionale di Digital Humanities. Lo sviluppo dell'applicazione prevede l'utilizzo e quindi la conoscenza avanzata di linguaggi di programmazione, framework e linguaggi per lo sviluppo web, protocolli e tecnologie per l'interrogazione e lo scambio di dati sul web e tecnologie del Semantic Web.