

Titolo:

La valorizzazione dei dati culturali attraverso metodologie semantiche. Il ruolo delle Digital Humanities (*Semantic methodologies for enhancing cultural data. The Digital Humanities role*)

Progetto di ricerca:

La ricerca si inserisce all'interno delle attività previste nel progetto "IDEHA - INNOVAZIONI PER L'ELABORAZIONE DEI DATI NEL SETTORE DEL PATRIMONIO CULTURALE" - codice identificativo ARS01_00421 - area di specializzazione "CULTURAL HERITAGE", finanziato nell'ambito dell'"Avviso per la presentazione di progetti di Ricerca Industriale e Sviluppo Sperimentale nelle 12 aree di specializzazione individuate dal PNR 2015 - 2020" (D.D. prot.1735 del 13 Luglio 2017).

In particolare la ricerca è relativa ai tasks di OR1, relativo alla costruzione di una piattaforma per ospitare dati culturali. In particolare, OR1 mira al tema della: "catalogazione, indicizzazione e descrizione delle risorse". La premessa teorica dell'intero progetto IDEHA origina dal fatto che:

L'ecosistema delle risorse digitali che a vario titolo riguardano il patrimonio culturale è ampio e molto ricco, ma soffre di un livello di frammentazione talmente alto da comprometterne il valore: i contenuti di qualità – prodotti da biblioteche, archivi e centri di ricerca - oltre a risultare spesso difficilmente reperibili e scarsamente accessibili, sono anche caratterizzati da un livello molto basso di integrazione, giacché questi soggetti lavorano spesso in contesti operativi e scientifici molto specifici, quando non in totale isolamento.

OR1 di IDEHA intende studiare e realizzare un catalogo delle risorse (Registry) che censisca, indicizzi e renda ricercabili le diverse entità: contenuti (dataset di ricerca, modelli 3D, file di modello d'informazioni), strumenti e servizi (BIM, Digital libraries, Visualizzatori etc.), provider (istituti, ricercatori etc.) ed altri elementi rilevanti nell'economia del progetto. Una piattaforma per la raccolta e aggregazione di contenuti digitali di ricerca scientificamente curati (dataset, authority list, corpora testuali etc.), riguardanti diversi aspetti del patrimonio culturale: elementi tangibili, intangibili e documentali.

L'obiettivo della presente ricerca è dunque contribuire alla produzione di dati culturali consistenti, secondo le specifiche del Semantic Web, che siano disseminati nella forma LOD (Linked Open Data) e che siano utili a popolare la knowledge base di IDEHA. In particolare il ruolo delle Digital Humanities è fondamentale nell'analisi dei modelli, nella ri-conciliazione e nella realizzazione di dati espressivi e consistenti.

La ricerca in OR1 si basa sui seguenti step progettuali:

- A1.2 Definizione delle strategie di acquisizione delle informazioni. Si intende analizzare e risolvere il problema della acquisizione di informazioni (dati e metadati) da diverse fonti, relativamente alle entità rilevanti per il progetto. In particolare si intendono sviluppare

soluzioni innovative per il data-mining e il web-crawling basate sull'analisi delle preferenze degli utenti.

- A1.3 Definizione delle strategie di mapping delle informazioni. Si intendono analizzare i differenti formati e standard di rappresentazione delle informazioni coinvolti nel progetto e si vogliono risolvere le questioni relative alla loro mappatura sintattica e semantica.
- A1.4 Definizione delle strategie per la modellazione delle informazioni. Si vogliono elaborare le soluzioni più efficaci per la rappresentazione delle informazioni in formato adatto all'interrogazione, interpretazione ed elaborazione automatica.
- A1.5 Definizione delle strategie per l'estrazione della conoscenza. Si intendono valutare diverse strategie ed approcci per l'estrazione della conoscenza dalle informazioni presenti nel sistema e la loro integrazione nella base di conoscenza integrata.
- A1.6 Definizione delle strategie per l'esposizione e l'accessibilità delle informazioni. Si vogliono mettere a punto le scelte per favorire l'accessibilità e l'interoperabilità delle informazioni, comprese le interfacce verso altre piattaforme, sistemi e infrastrutture.

La ricerca si inserisce in questo workflow cercando di stabilire come le Digital Humanities possano contribuire ai diversi step, con un'attenzione speciale ai dati in formato LOD, al problema della possibile coesistenza di punti di vista concorrenti su quei dati e al tema dei dati provenienti dal full-text delle edizioni digitali di testi, prevenendo la possibilità di un loro riuso in vista dell'arricchimento dei dati culturali presenti nel Registry.

Speciale attenzione verrà riservata, nella ricerca, al dialogo di OR1 con OR7 su "Modelli di business e tecnologie per la valorizzazione dei beni culturali", in particolare per gli aspetti concernenti la creazione di musei virtuali e la gestione di itinerari tematici smart.

Piano di attività:

Per ottemperare a questi scopi le attività previste dal progetto riguardano:

- Analisi del Registry dei dati culturali;
- Analisi delle esistenti ontologie per il patrimonio culturale (es. CIDOC-CRM, EAD/EAC, FRBR) e dei vocabolari per la metadattazione (es. DC, TEI, METS, MODS) nonché dei vocabolari controllati (es. Getty Thesaurus, WordNet, Geonames, Dbpedia, etc..) e dei sistemi di soggettazione e classificazione in uso nella catalogazione, descrizione, indicizzazione e classificazione dei dati dei beni culturali;
- Analisi delle edizioni digitali già presenti in rete. Valutazione dello stato dell'arte;
- Progettazione di un modello di gestione di posizioni concorrenti sull'interpretazione dei dati (possibile studio dell'apparato delle edizioni critico come modello per la gestione di punti di vista eterogenei);
- Modellazione del processo di creazione di dati espressivi dal full-text delle edizioni;
- Proposta di arricchimento dei LOD esistenti attraverso dati provenienti da altre fonti;
- Creazione di strumenti di valorizzazione dell'accesso alle risorse culturali così realizzate.