

Specifiche per l'esame di Modellazione concettuale per il Web Semantico - 2024/25

Per chi fa laboratorio con SPARQLanything

- Singoli: non fanno Query SPARQL e non fanno estensione
- Gruppi
 - Non fanno Query SPARQL
 - Fanno una sola estensione (a scelta tra 2,3,4,5,6)
- Il progetto vale 70%

Parte II

Basandosi sull'ontologia creata per la prima parte del progetto (vedi documento di specifiche), realizzare un insieme di query e una o più estensioni (vedere sotto):

1. Queries SPARQL

Tutti (singoli e gruppi di due persone):

Progettare un flusso di interazione con l'utente che sia coerente con i requirements dichiarati nel documento di progettazione dell'ontologia (tipo di utenti, task, ecc.). Allegare:

- flow chart dell'interazione
- schema di interfaccia (mockup senza grafica)
- esempio di interazione con dati reali

Creare le query (minimo 5) a supporto dell'interazione progettata che permettono di estrarre le informazioni contenute nel knowledge graph. Per ogni query, inserire:

- la query stessa
- descrizione della query
- risultato sui dati di esempio

Le query devono includere almeno un COUNT

Gruppi di due persone

(oltre alla parte precedente, minimo 8 queries) Inserire almeno 2 queries che interrogano un endpoint SPARQL pubblico (es. dbpedia, wikidata, VIAF, ecc.) per integrare i dati estratti dall'ontologia (per esempio, con informazioni di tipo localizzazione, di autorità, ecc.).

È possibile creare query federate oppure integrare i dati per via programmatica.

2. Estensioni:

Progetti singoli:

Una estensione a scelta (con l'eccezione di chi fa laboratorio SPARQLanything).

Gruppi di due persone:

Applicazione client e una estensione a scelta tra le restanti (con l'eccezione di chi fa laboratorio SPARQLanything).

1. Applicazione client

- Utilizzando una Linked Data Platform (Virtuoso, GraphDB, BlazeGraph o altri – incluse librerie per data storage e manipulation come Jena) e il suo SPARQL endpoint (o API fornita dalla LDP), creare l'applicazione client che interroga l'ontologia secondo lo schema definito in precedenza (via SPARQL o API). Se la LDP lo supporta, è possibile creare una o più query per l'inserimento di dati.
- Per i gruppi di due persone, includere l'interrogazione di altre sorgenti di dati.
- Note:
 - La grafica dell'interfaccia non è rilevante
 - Il codice deve essere commentato.
 - Preparare una documentazione sintetica del progetto (1/2 pagina)

2. Importazione dei dati nel grafo SPARQLanything

Utilizzando un plugin per l'importazione dei dati da schema relazionale a RDF via template R2RML (OnTop o Mastro OBDA), realizzare l'importazione dei dati nel knowledge graph a partire da un database esistente (o un prototipo creato ad hoc).

Descrivere sinteticamente l'esportazione (1 pagina) e allegare:

- Lo schema del database
- Il mapping in formato R2RML (con commenti)
- Le triple materializzate

3. Importazione dei dati nel grafo

Utilizzando un plugin per l'importazione dei dati da schema relazionale a RDF via template R2RML (OnTop o Mastro OBDA), realizzare l'importazione dei dati nel knowledge graph a partire da un database esistente (o un prototipo creato ad hoc).

Descrivere sinteticamente l'esportazione (1 pagina) e allegare:

- Lo schema del database
- Il mapping in formato R2RML (con commenti)
- Le triple materializzate

4. Base di regole SWRL

- Creare una base di 5 regole SWRL che aggiungono nuove asserzioni sugli individui (classi e proprietà) o mappano gli individui di due ontologie tra di loro (non è necessario che l'ontologia con cui si realizza il mapping via SWRL sia quella utilizzata per l'integrazione nella fase di sviluppo della propria ontologia).
- Ad ogni regola associare un commento che ne descrive la funzione.
- Almeno una regola deve contenere un operatore built-in.

5. Creazione automatica di ontologie (lezioni Prof. Daga)

Utilizzando tecniche di NLP (da Python NLTK, AMRLib, Spacey, ...) e tools per l'estrazione di entità (e.g. DBpediaSpotlight, FRED, Relik), generare una bozza di ontologia che include Classi, Proprietà a partire da una collezione di competency questions (CQ).

- (Si può provare a tentare di estrarre automaticamente relazioni tra concetti: subPropertyOf, subClassOf, domain, range). Partire da questo esercizio come esempio: Ontology drafting from Competency Questions (Colab) in <https://github.com/enridaga/turin24/> (-> December)
- Si può usare questo dataset come fonte di competency questions e ontologie di riferimento: Potoniec, Jędrzej; Wiśniewski, Dawid ; Ławrynowicz, Agnieszka; Keet, Maria (2019), "Ontology Competency Questions to SPARQL-OWL Queries Translations", Mendeley Data, V1, doi: 10.17632/pp6hmfxfgf.1 <https://data.mendeley.com/datasets/pp6hmfxfgf/1>

Si può usare un large language model (LLM) MA solo tramite API all'interno di un processo interamente automatico, in uno script, o Notebook, Python.

Nella relazione (ma anche nei commenti al codice) si discutano le seguenti problematiche:

- "Come estrarre automaticamente concetti e relazioni dal linguaggio naturale?"
- "Cosa rende una parola rilevante o meno per l'ontologia?"

- “Che metodi possono aiutare per escludere termini che non sono rilevanti per l’ontologia?”
- “Come migliorare la qualità delle entità generate, considerando le buone pratiche in Ontology Engineering?”

6. Estensioni a scelta

I gruppi o i singoli che lo desiderano possono sviluppare altre estensioni usando strumenti avanzati o sperimentali, se lo desiderano ([RDF*](#), [Omeka-S](#), [suite commerciali](#), ecc.): in questo caso, contattatemi con un po’ di anticipo per discuterne.