

Progetto IA2 (a.a. 2016/2017)

Il progetto consiste nello sviluppo di un sistema di *question answering per basi di conoscenza* aderenti agli standard di rappresentazione usati nel contesto del Semantic Web.

Il sistema deve accettare domande formulate in lingua Inglese, tradurle nel linguaggio (formale) di interrogazione SPARQL e sottometterle ad una base di conoscenza per generare le risposte.

Le prove saranno effettuate su una specifica base di conoscenza, costruita durante lo sviluppo del sistema. Ciononostante, il progetto del sistema deve supportare il suo adattamento ad ontologie differenti.

Il progetto può essere svolto in gruppi formati al massimo da tre persone. Ciascun gruppo dovrebbe scegliere un dominio diverso, o quantomeno approfondire un aspetto diverso di uno stesso dominio. La scelta del dominio e la formulazione delle domande sono affidate al gruppo stesso. In fondo al documento, tuttavia, sono state predisposte alcune domande (relative a due domini diversi) che sono esemplificative del grado di supporto che ci si aspetta dal sistema sviluppato.

L'esecuzione del progetto deve essere basata sull'approccio, le tecniche e le tecnologie che saranno discusse nell'ambito del corso.

Viene suggerita la seguente scaletta di attività:

- Analisi del dominio e dei requisiti
 - Formulazione delle *domande*
 - Enumerazione dei *termini rilevanti*
- Costruzione di una ontologia di dominio
 - Informata dalle *domande*, da usare anche per la *valutazione dell'ontologia* (es. l'ontologia è in grado di rispondere alle *domande*?)
- Definizione di un lessico in lingua Inglese per l'ontologia di dominio (rappresentato secondo il modello Lemon)
 - La costruzione del lessico può essere guidata dalle *domande* identificate in precedenza
- Produzione grammatica LTAG
 - Scrittura della *parte indipendente dal dominio*
 - Generazione della *parte specifica del dominio* dal lessico sviluppato in precedenza
- Uso di un parser LTAG per analisi delle domande e loro trasformazione in SPARQL

Il processo di cui sopra dovrebbe essere eseguito in maniera iterativa, considerando nelle iterazioni successive delle *domande addizionali* che coprano aspetti quali:

- composizionalità,
- varianti lessicali (sinonimi)
- diversi costrutti sintattici (es. l'uso del genitivo sassone al posto della preposizione "of")

Extra:

Extra 1. Considerare alberi ausiliari.

Extra 2. Sfruttare l'ontologia sottostante per risolvere fenomeni di ambiguità o vaghezza

Al termine del progetto, sarà necessario consegnare con debito anticipo:

- i sorgenti (accompagnati da istruzioni per compilare ed eseguire il sistema)
- una relazione (del lavoro svolto)

Nel corso della prova orale, sarà inoltre richiesto di mostrare e discutere

- una presentazione sul progetto realizzato

DOMINIO 1 - ORGANIZATIONS

1. Who founded Microsoft? (active form of a transitive verb)
2. Who are the founders of Microsoft? (relational noun)
3. How many people founded Microsoft? (counting + class noun [people])
4. Who is the CEO of Apple? (relational noun)
5. What is the name of the CEO of Apple? (nested relational nouns)
6. Who is the chief executive officer of Apple? (relational noun + multi-word expression + full form of acronym)
7. Who is the chief financial officer of Apple? (relational noun + multi-word expression)
8. Who are the corporate officers of Apple? (corporate officer subsumes CEO and CFO according to WordNet)
9. Who is the chairman of Apple? (relational noun)
10. Who is the president of Google? (relational noun + a synonym of chairman according to WordNet)
11. What is the net income of Microsoft? (multiword expression + data property in the ontology)

Alcuni esempi di domande “extra”:

- Extra 1. Is Satya Nadella the CEO of Microsoft? (yes/no question)
- Extra 2. Did Microsoft acquire a company headquartered in Italy? (yes/no question + relational adjective)
- Extra 3. Did Microsoft acquire an Italian company? (yes/no question + adjective [it should mean headquartered in Italy])
- Extra 4. Is Satya Nadella Italian? (predicative adjective, ambiguity wrt to the use in Italian company)
- Extra 5. Where is Microsoft headquartered? (where + relational adjective)
- Extra 6. What is the most valuable company? (superlative)

DOMINIO 2 – ARTWORKS

1. Who painted Mona Lisa? (active form of a transitive verb)
2. Who is the painter of Mona Lisa? (relational noun)
3. Which museum conserves Mona Lisa? (active verb)
4. How many statues did Michelangelo create? (counting + class noun [people])
5. Who is the painter of the Last Judgment? (relational noun)
6. Who is the name of the painter of the Last Judgment? (nested relational nouns)
7. Who is the artist of the Last Judgment? (relational noun + artist subsumes painter according to WordNet)
8. What is the height of the David? (relational noun + data property)

Alcuni esempi di domande “extra”:

- Extra 1. Is Michelangelo the author of the David? (yes/no question)
- Extra 2. Which works were painted by David? (passive, ambiguity between the statue and the painter)
- Extra 3. Is the David made of bronze? (yes/no question + relational adjective)
- Extra 4. Is the David a bronze statue? (yes/no question + adjective [it should mean made of bronze])
- Extra 5. Where is Mona Lisa? (where)
- Extra 6. What is the highest statue? (superlative)