

Introduzione ai Linked Open Data e al Web Semantico

Cristiano Longo
longo@dmf.unict.it

Università di Catania, 11 Giugno 2015

Questa presentazione tratterà i seguenti argomenti:

- Motivazioni del Web Semantico
- Definizione formale di Ontologie
- Interrogazioni sulle Ontologie
- Vocabolari

Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).¹

da *Open Knowledge Foundation: The Open Definition*

¹<http://opendefinition.org>

Nelle *LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO*, emanate dall'*Agenzia per l'Italia Digitale (AgID)*

vengono riportate le seguenti definizioni

- a) *formato dei dati di tipo aperto*, un formato di dati reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi;

- b) *dati di tipo aperto*, i dati che presentano le seguenti caratteristiche: 1) sono disponibili secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato; 2) sono accessibili attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti ai sensi della lettera a), sono adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi *metadati*;

- 3) sono resi disponibili gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione.

L'Agenzia per l'Italia digitale deve stabilire, con propria deliberazione, i casi eccezionali, individuati secondo criteri oggettivi, trasparenti e verificabili, in cui essi sono resi disponibili a tariffe superiori ai costi marginali. [...]

Nelle *LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO*, emanate dall'*Agenzia per l'Italia Digitale (AgID)* vengono riportate le seguenti definizioni a) *formato dei dati di tipo aperto*, un formato di dati reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi;

b) *dati di tipo aperto*, i dati che presentano le seguenti caratteristiche: 1) sono disponibili secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato; 2) sono accessibili attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti ai sensi della lettera a), sono adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi *metadati*;

3) sono resi disponibili gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione.

L'Agenzia per l'Italia digitale deve stabilire, con propria deliberazione, i casi eccezionali, individuati secondo criteri oggettivi, trasparenti e verificabili, in cui essi sono resi disponibili a tariffe superiori ai costi marginali. [...]

Definizione di Open Data - Agenzia per l'Italia Digitale

Nelle *LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO*, emanate dall'*Agenzia per l'Italia Digitale (AgID)* vengono riportate le seguenti definizioni a) *formato dei dati di tipo aperto*, un formato di dati reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi;

b) *dati di tipo aperto*, i dati che presentano le seguenti caratteristiche: 1) sono disponibili secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato; 2) sono accessibili attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti ai sensi della lettera a), sono adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi *metadati*;

3) sono resi disponibili gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione.

L'Agenzia per l'Italia digitale deve stabilire, con propria deliberazione, i casi eccezionali, individuati secondo criteri oggettivi, trasparenti e verificabili, in cui essi sono resi disponibili a tariffe superiori ai costi marginali. [...]

Definizione di Open Data - Agenzia per l'Italia Digitale

Nelle *LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO*, emanate dall'*Agenzia per l'Italia Digitale (AgID)* vengono riportate le seguenti definizioni a) *formato dei dati di tipo aperto*, un formato di dati reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi;

b) *dati di tipo aperto*, i dati che presentano le seguenti caratteristiche: 1) sono disponibili secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato; 2) sono accessibili attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti ai sensi della lettera a), sono adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi *metadati*;

3) sono resi disponibili gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione.

L'Agenzia per l'Italia digitale deve stabilire, con propria deliberazione, i casi eccezionali, individuati secondo criteri oggettivi, trasparenti e verificabili, in cui essi sono resi disponibili a tariffe superiori ai costi marginali. [...]

Definizione di Open Data - Agenzia per l'Italia Digitale

Nelle *LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO*, emanate dall'*Agenzia per l'Italia Digitale (AgID)* vengono riportate le seguenti definizioni a) *formato dei dati di tipo aperto*, un formato di dati reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi;

b) *dati di tipo aperto*, i dati che presentano le seguenti caratteristiche: 1) sono disponibili secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato; 2) sono accessibili attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti ai sensi della lettera a), sono adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi *metadati*;

3) sono resi disponibili gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione.

L'Agenzia per l'Italia digitale deve stabilire, con propria deliberazione, i casi eccezionali, individuati secondo criteri oggettivi, trasparenti e verificabili, in cui essi sono resi disponibili a tariffe superiori ai costi marginali. [...]

Definizione di Open Data - Agenzia per l'Italia Digitale

Nelle *LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO*, emanate dall'*Agenzia per l'Italia Digitale (AgID)* vengono riportate le seguenti definizioni a) *formato dei dati di tipo aperto*, un formato di dati reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi;

b) *dati di tipo aperto*, i dati che presentano le seguenti caratteristiche: 1) sono disponibili secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato; 2) sono accessibili attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti ai sensi della lettera a), sono adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi *metadati*;

3) sono resi disponibili gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione.

L'Agenzia per l'Italia digitale deve stabilire, con propria deliberazione, i casi eccezionali, individuati secondo criteri oggettivi, trasparenti e verificabili, in cui essi sono resi disponibili a tariffe superiori ai costi marginali. [...]

Definizione di Open Data - Agenzia per l'Italia Digitale

Nelle *LINEE GUIDA NAZIONALI PER LA VALORIZZAZIONE DEL PATRIMONIO INFORMATIVO PUBBLICO*, emanate dall'*Agenzia per l'Italia Digitale (AgID)* vengono riportate le seguenti definizioni a) *formato dei dati di tipo aperto*, un formato di dati reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi;

b) *dati di tipo aperto*, i dati che presentano le seguenti caratteristiche: 1) sono disponibili secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato; 2) sono accessibili attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti ai sensi della lettera a), sono adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi *metadati*;

3) sono resi disponibili gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione. L'Agenzia per l'Italia digitale deve stabilire, con propria deliberazione, i casi eccezionali, individuati secondo criteri oggettivi, trasparenti e verificabili, in cui essi sono resi disponibili a tariffe superiori ai costi marginali. [...]

Tecnicamente, è sufficiente pubblicare le informazioni su un sito web o come documenti *non strutturati*, ma usando formati aperti?

The Web was designed as an information space, with the goal that it should be useful not only for human-human communication, but also that machines would be able to participate and help. One of the major obstacles to this has been the fact that most information on the Web is designed for human consumption, and even if it was derived from a database with well defined meanings (in at least some terms) for its columns, that the structure of the data is not evident to a robot browsing the web.

Semantic Web Roadmap, Tim Berners-Lee, 1998.

Tecnicamente, è sufficiente pubblicare le informazioni su un sito web o come documenti *non strutturati*, ma usando formati aperti?

*The Web was designed as an information space, with the goal that it should be useful not only for human-human communication, but also that machines would be able to participate and help. One of the major obstacles to this has been the fact **that most information on the Web is designed for human consumption**, and even if it was derived from a database with well defined meanings (in at least some terms) for its columns, that **the structure of the data is not evident to a robot browsing the web.***

Semantic Web Roadmap, Tim Berners-Lee, 1998.

Alcuni problemi nell'interpretazione di testi derivano da:

Lingue Differenti e.g. *Parigi* e *Paris* possono indicare la stessa città.

Omonimie e.g. esistono svariate città chiamate *Paris* nel mondo (Arkansas, Idaho, Illinois, Kentucky, Maine, Michigan, Missouri, New York, ...);

Limiti del World Wide Web (3/4)

La situazione si complica in presenza di contenuti multimediali.



Come conseguenza, spesso è impossibile eseguire su web ricerche *complesse* ottenendo risultati accurati. Ad esempio, cercando sul web "*Federico II places*" non si ottengono risultati in prima pagina su Federico II, ma solo sull'omonima università:

- ① Università degli Studi di Napoli "Federico II" — OPEN Places
- ② AOU - Policlinico "Federico II" - Napoli, Italy - Hospital — Facebook
- ③ Federico II Ingegneria Via Claudio - College and University — Facebook
- ④ MARIA CATERINA FONTE - www.docenti.unina.it

II Web Semantico

[...] the Semantic Web approach instead develops languages for expressing information in a machine processable form.

Semantic Web Roadmap, Tim Berners-Lee, 1998.

About: [Frederick II, Holy Roman Emperor](#) [About DBpedia](#)

An Entity of Type : [agent](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

Frederick II (26 December 1194 – 13 December 1250), was one of the most powerful Holy Roman Emperors of the Middle Ages and head of the House of Hohenstaufen. His political and cultural ambitions, based in Sicily and stretching through Italy to Germany, and even to Jerusalem, were enormous; however, his enemies, especially the popes, prevailed, and his dynasty collapsed soon after his death.

Property	Value
dbpedia-owl:abstract	■ 1250-01-01 00:00:00 (xsd:date)
dbpedia-owl:activeYearsStartYear	■ 1220-01-01 00:00:00 (xsd:date)
dbpedia-owl:birthDate	■ 1194-12-26 (xsd:date)
dbpedia-owl:birthPlace	■ dbpedia:lesi ■ dbpedia:Kingdom_of_Italy_(medieval) ■ dbpedia:Marche
dbpedia-owl:deathDate	■ 1250-12-13 (xsd:date)
dbpedia-owl:deathPlace	■ dbpedia:Torremaqqiore

Figure : Federico II su dbpedia.org

Linked Open Data Cloud (1/2)

The Semantic Web is a web of data, in some ways like a global database.

Semantic Web Roadmap, Tim Berners-Lee, 1998.

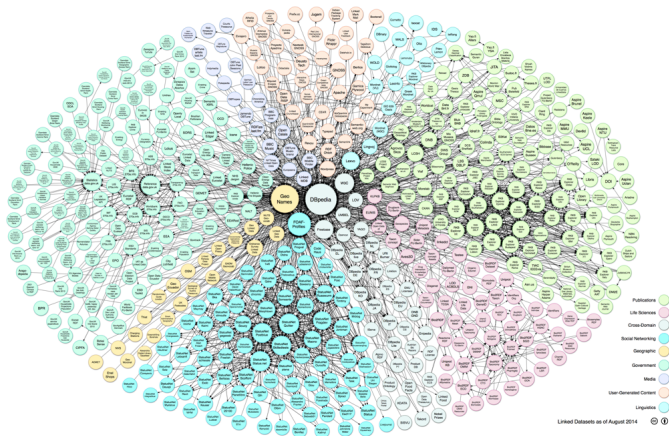


Figure : Linked Open Data Cloud

La classificazione 5 stelle viene estesa dall'AgID con alcune *dimensioni* esplicative:

- **INFORMAZIONE** - descrive la qualità dell'informazione fornita insieme ai dati;
- **ACCESSO** - descrive la facilità con cui utenti e programmi riescono ad accedere ai dati;
- **SERVIZI** - riguarda le tipologie e l'*efficienza* dei servizi che possono essere realizzati a partire dai dati.

La classificazione 5 stelle viene estesa dall'AgID con alcune *dimensioni* esplicative:

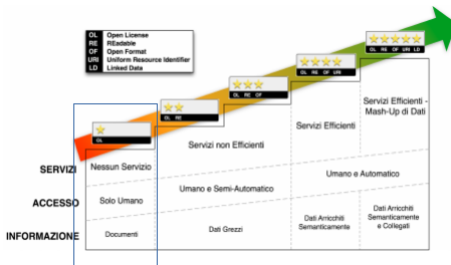
- **INFORMAZIONE** - descrive la qualità dell'informazione fornita insieme ai dati;
- **ACCESSO** - descrive la facilità con cui utenti e programmi riescono ad accedere ai dati;
- **SERVIZI** - riguarda le tipologie e l'*efficienza* dei servizi che possono essere realizzati a partire dai dati.

La classificazione 5 stelle viene estesa dall'AgID con alcune *dimensioni* esplicative:

- *INFORMAZIONE* - descrive la qualità dell'informazione fornita insieme ai dati;
- *ACCESSO* - descrive la facilità con cui utenti e programmi riescono ad accedere ai dati;
- *SERVIZI* - riguarda le tipologie e l'*efficienza* dei servizi che possono essere realizzati a partire dai dati.

Classificazione 5 Stelle - Una Stella

Una Stella: Open Licence

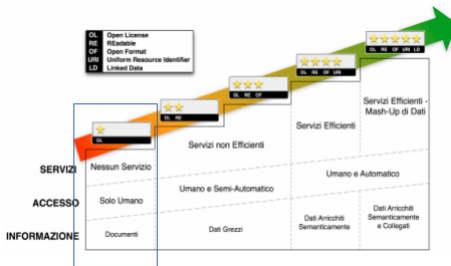


La *prima stella* si ottiene rilasciando i dati in qualunque formato ma con una *licenza aperta*. Rientrano in questa categoria ad esempio le scansioni dei documenti.

- **INFORMAZIONE:** *documenti* - i dati sono incorporati all'interno di documenti senza struttura;
- **ACCESSO:** *solo umano* - solo gli umani sono in grado di leggere i documenti senza struttura e quindi dare un senso ai dati in esso presenti;
- **SERVIZI:** *nessuno* può essere abilitato a meno di significativi interventi umani di estrazione ed elaborazione;

Classificazione 5 Stelle - Una Stella

Una Stella: Open Licence

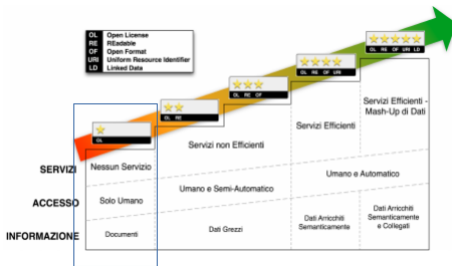


La *prima stella* si ottiene rilasciando i dati in qualunque formato ma con una *licenza aperta*. Rientrano in questa categoria ad esempio le scansioni dei documenti.

- **INFORMAZIONE:** *documenti* - i dati sono incorporati all'interno di documenti senza struttura;
- **ACCESSO:** *solo umano* - solo gli umani sono in grado di leggere i documenti senza struttura e quindi dare un senso ai dati in esso presenti;
- **SERVIZI:** *nessuno* può essere abilitato a meno di significativi interventi umani di estrazione ed elaborazione;

Classificazione 5 Stelle - Una Stella

Una Stella: Open Licence

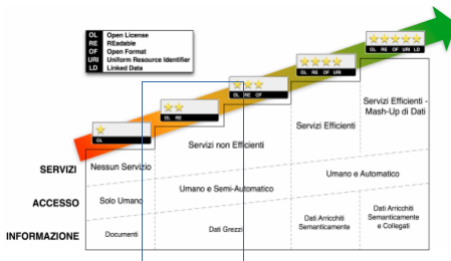


La *prima stella* si ottiene rilasciando i dati in qualunque formato ma con una *licenza aperta*. Rientrano in questa categoria ad esempio le scansioni dei documenti.

- **INFORMAZIONE:** *documenti* - i dati sono incorporati all'interno di documenti senza struttura;
- **ACCESSO:** *solo umano* - solo gli umani sono in grado di leggere i documenti senza struttura e quindi dare un senso ai dati in esso presenti;
- **SERVIZI:** *nessuno* può essere abilitato a meno di significativi interventi umani di estrazione ed elaborazione;

Classificazione 5 Stelle - Due Stelle

Due Stelle: Open Licence, (Machine) Readable



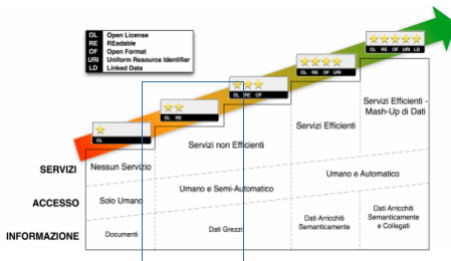
La *seconda stella* si ottiene se i dati sono forniti in un formato leggibile da un agente automatico.

Rientrano in questa categoria ad esempio i files in formato *excel*.

- **INFORMAZIONE:** *dati grezzi (o semi-strutturati)* - i dati sono leggibili anche da un programma ma necessita un intervento umano per interpretarli;
- **ACCESSO:** *umano e semi-automatico* - i software possono leggere i dati ma non sono in grado di interpretarli automaticamente;
- **SERVIZI:** *non efficienti* - servizi realizzati ad-hoc e devono incorporare al loro interno i dati;

Classificazione 5 Stelle - Due Stelle

Due Stelle: Open Licence, (Machine) Readable



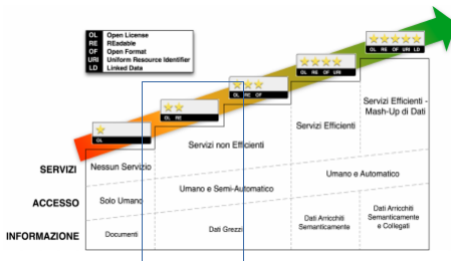
La *seconda stella* si ottiene se i dati sono forniti in un formato leggibile da un agente automatico.

Rientrano in questa categoria ad esempio i files in formato *excel*.

- **INFORMAZIONE:** *dati grezzi (o semi-strutturati)* - i dati sono leggibili anche da un programma ma necessita un intervento umano per interpretarli;
- **ACCESSO:** *umano e semi-automatico* - i software possono leggere i dati ma non sono in grado di interpretarli automaticamente;
- **SERVIZI:** *non efficienti* - servizi realizzati ad-hoc e devono incorporare al loro interno i dati;

Classificazione 5 Stelle - Due Stelle

Due Stelle: Open Licence, (Machine) Readable



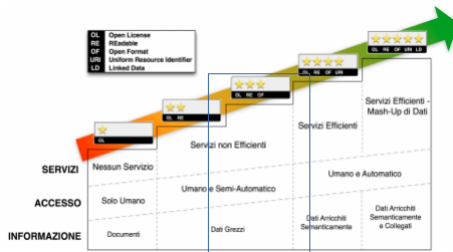
La *seconda stella* si ottiene se i dati sono forniti in un formato leggibile da un agente automatico.

Rientrano in questa categoria ad esempio i files in formato *excel*.

- **INFORMAZIONE:** *dati grezzi (o semi-strutturati)* - i dati sono leggibili anche da un programma ma necessita un intervento umano per interpretarli;
- **ACCESSO:** *umano e semi-automatico* - i software possono leggere i dati ma non sono in grado di interpretarli automaticamente;
- **SERVIZI:** *non efficienti* - servizi realizzati ad-hoc e devono incorporare al loro interno i dati;

Classificazione 5 Stelle - Tre Stelle

Tre Stelle: Open Licence, (Machine) Readable, Open Format

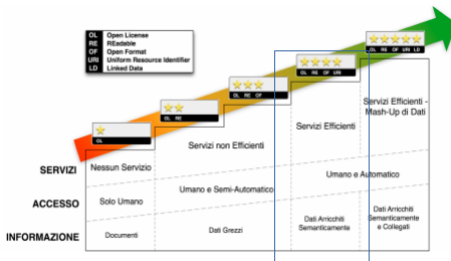


La *terza stella* viene attribuita se i dati sono rilasciati in un formato *aperto*. Rientrano in questa categoria ad esempio i files *json*, *csv*, *xml*.

- **INFORMAZIONE:** *dati grezzi (o semi-strutturati)* - i dati sono leggibili anche da un programma ma necessita un intervento umano per interpretarli;
- **ACCESSO:** *umano e semi-automatico* - i software possono leggere i dati ma non sono in grado di interpretarli automaticamente;
- **SERVIZI:** *non efficienti* - servizi realizzati ad-hoc e devono incorporare al loro interno i dati;

Classificazione 5 Stelle - Quattro Stelle

Quattro Stelle: Open Licence, (Machine) Readable, Open Format, URI

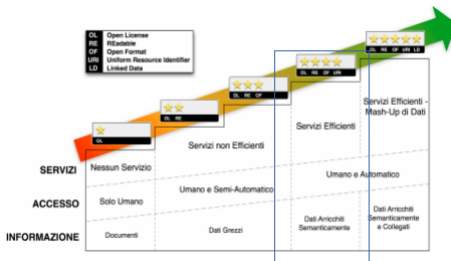


La *quarta stella* si ottiene esponendo i dati con le tecnologie del web semantico (RDF e SPARQL).

- **INFORMAZIONE:** *dati arricchiti semanticamente* - i dati sono descritti usando tecnologie del Web Semantico;
- **ACCESSO:** *umano e automatico* - i software sono in grado di elaborare i dati quasi senza ulteriori interventi umani (livelli 4 e 5);
- **SERVIZI:** *efficienti* - servizi che sfruttano accessi diretti a Web per reperire i dati.

Classificazione 5 Stelle - Quattro Stelle

Quattro Stelle: Open Licence, (Machine) Readable, Open Format, URI

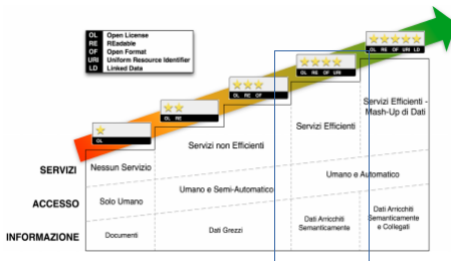


La *quarta stella* si ottiene esponendo i dati con le tecnologie del web semantico (RDF e SPARQL).

- **INFORMAZIONE:** *dati arricchiti semanticamente* - i dati sono descritti usando tecnologie del Web Semantico;
- **ACCESSO:** *umano e automatico* - i software sono in grado di elaborare i dati quasi senza ulteriori interventi umani (livelli 4 e 5);
- **SERVIZI:** *efficienti* - servizi che sfruttano accessi diretti a Web per reperire i dati.

Classificazione 5 Stelle - Quattro Stelle

Quattro Stelle: Open Licence, (Machine) Readable, Open Format, URI

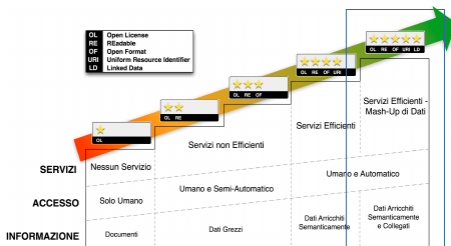


La *quarta stella* si ottiene esponendo i dati con le tecnologie del web semantico (RDF e SPARQL).

- **INFORMAZIONE:** *dati arricchiti semanticamente* - i dati sono descritti usando tecnologie del Web Semantico;
- **ACCESSO:** *umano e automatico* - i software sono in grado di elaborare i dati quasi senza ulteriori interventi umani (livelli 4 e 5);
- **SERVIZI:** *efficienti* - servizi che sfruttano accessi diretti a Web per reperire i dati.

Classificazione 5 Stelle - Cinque Stelle

Cinque Stelle: Open Licence, (Machine) Readable, Open Format, URI, *Linked* Data

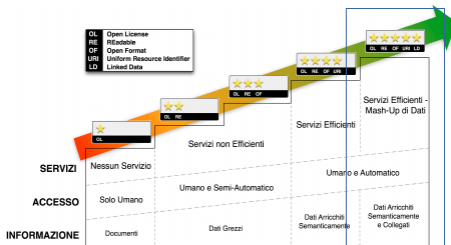


La *quinta stella* viene attribuita quando i dati contengono riferimenti a dataset di terze parti.

- **INFORMAZIONE:** *dati arricchiti semanticamente* - i dati sono descritti usando tecnologie del Web Semantico;
- **ACCESSO:** *umano e automatico* - i software sono in grado di elaborare i dati quasi senza ulteriori interventi umani (livelli 4 e 5);
- **SERVIZI:** *efficienti e con mashup di dati* - servizi che sfruttano sia accessi diretti a Web sia l'informazione ulteriore catturata attraverso i *link* dei dati di interesse.

Classificazione 5 Stelle - Cinque Stelle

Cinque Stelle: Open Licence, (Machine) Readable, Open Format, URI, *Linked* Data

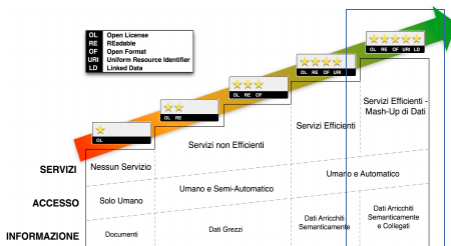


La *quinta stella* viene attribuita quando i dati contengono riferimenti a dataset di terze parti.

- **INFORMAZIONE:** *dati arricchiti semanticamente* - i dati sono descritti usando tecnologie del Web Semantico;
- **ACCESSO:** *umano e automatico* - i software sono in grado di elaborare i dati quasi senza ulteriori interventi umani (livelli 4 e 5);
- **SERVIZI:** *efficienti e con mashup di dati* - servizi che sfruttano sia accessi diretti a Web sia l'informazione ulteriore catturata attraverso i *link* dei dati di interesse.

Classificazione 5 Stelle - Cinque Stelle

Cinque Stelle: Open Licence, (Machine) Readable, Open Format, URI, *Linked* Data



La *quinta stella* viene attribuita quando i dati contengono riferimenti a dataset di terze parti.

- **INFORMAZIONE:** *dati arricchiti semanticamente* - i dati sono descritti usando tecnologie del Web Semantico;
- **ACCESSO:** *umano e automatico* - i software sono in grado di elaborare i dati quasi senza ulteriori interventi umani (livelli 4 e 5);
- **SERVIZI:** *efficienti e con mashup di dati* - servizi che sfruttano sia accessi diretti a Web sia l'informazione ulteriore catturata attraverso i *link* dei dati di interesse.

I dataset del Web Semantico vengono spesso definiti *ontologie*.

Una *ontologia* è una descrizione *parziale* del mondo:

- descrive una porzione del mondo, spesso è limitata ad un'unico *dominio di conoscenza*;
- non si assume che i fatti non esplicitamente presenti nell'ontologia siano falsi (*Open World Assumption*).

Essa è costituita da un insieme finito di *affermazioni*. Ad esempio:

- Tutti gli esseri umani sono mortali;
- Socrate è mortale;
- Alice è la madre di Roberto.

Le affermazioni contenute in una ontologia sono di tre tipi:

Constraints: impongono dei vincoli *semantici* sul dominio di conoscenza che si va a rappresentare. La notazione richiama quella insiemistica;

$$\textit{HumanBeing} \sqsubseteq \textit{Mortal}$$

Property Assertions: impongono una relazione tra due elementi del dominio;

$$\textit{Alice} \textit{ motherOf } \textit{Bob}$$

Class Assertions: indicano l'appartenenza di un elemento ad un insieme.

$$\textit{HumanBeing}(\textit{Socrate})$$

Con il termine *reasoning* si intende l'attività di estrazione di conoscenza *implicita* in una ontologia.

$$\left\{ \begin{array}{l} HumanBeing \sqsubseteq Mortal, \\ HumanBeing(Socrate) \end{array} \right\} \implies Mortal(Socrate)$$

Le attività di reasoning sono rese possibili dalle *semantiche formali* associate ai linguaggi di rappresentazione utilizzati.

Siano N_C , N_P , N_I tre insiemi infiniti, numerabili e a due a due disgiunti di nomi di classe, proprietà e individuo, rispettivamente.

Una *ontologia* è un insieme finito di asserzioni dei seguenti tipi:

	<i>Sintassi</i>	<i>Semantica</i>
Constraints	$C \sqsubseteq D$ $R \sqsubseteq S$ $\text{dom}(R) \sqsubseteq C$ $\text{range}(R) \sqsubseteq C$	$(\forall x)(x \in C \rightarrow x \in D)$ $(\forall x, y)([x, y] \in R \rightarrow [x, y] \in S)$ $(\forall x, y)([x, y] \in R \rightarrow x \in C)$ $(\forall x, y)([x, y] \in R \rightarrow y \in C)$
Class Assertions	$C(a)$	$a \in C$
Property Assertions	$a P b$ (equivalente $P(a, b)$)	$[a, b] \in P$

dove $C, D \in N_C$, $R, S \in N_P$ e $a, b \in N_I$.

Riportiamo un esempio di ontologia. Siano $HumanBeing, Mortal \in N_C$,
 $teacherOf \in N_P$, $Socrate, Platone \in N_I$.

$$\mathcal{O} = \{ HumanBeing \sqsubseteq Mortal, \\ range(teacherOf) \sqsubseteq HumanBeing, \\ HumanBeing(Socrate), \\ Socrate \text{ teacherOf } Platone \}$$

Riportiamo un esempio di ontologia. Siano $HumanBeing, Mortal \in N_C$,
 $teacherOf \in N_P$, $Socrate, Platone \in N_I$.

Mediante *reasoning* è possibile esplicitare ulteriori affermazioni.

$$\begin{aligned} \mathcal{O} = \{ & \mathbf{HumanBeing} \sqsubseteq \mathbf{Mortal}, \\ & \text{range}(teacherOf) \sqsubseteq \mathbf{HumanBeing}, \\ & \mathbf{HumanBeing}(\mathbf{Socrate}), \\ & \mathbf{Socrate} \text{ teacherOf } \mathbf{Platone} \} \end{aligned} \quad \Longrightarrow \quad \mathcal{O}' = \{ \mathbf{Mortal}(\mathbf{Socrate}), \}$$

Riportiamo un esempio di ontologia. Siano $HumanBeing, Mortal \in N_C$,
 $teacherOf \in N_P$, $Socrate, Platone \in N_I$.

Mediante *reasoning* è possibile esplicitare ulteriori affermazioni.

$$\begin{aligned} \mathcal{O} = \{ & HumanBeing \sqsubseteq Mortal, \\ & range(teacherOf) \sqsubseteq HumanBeing, \\ & HumanBeing(Socrate), \\ & Socrate teacherOf Platone \} \end{aligned} \quad \Rightarrow \quad \begin{aligned} \mathcal{O}' = \{ & Mortal(Socrate), \\ & HumanBeing(Platone), \\ & \} \end{aligned}$$

Riportiamo un esempio di ontologia. Siano $HumanBeing, Mortal \in N_C$,
 $teacherOf \in N_P$, $Socrate, Platone \in N_I$.

Mediante *reasoning* è possibile esplicitare ulteriori affermazioni.

$$\begin{aligned} \mathcal{O} = \{ & \mathbf{HumanBeing} \sqsubseteq \mathbf{Mortal} \\ & \text{range}(teacherOf) \sqsubseteq HumanBeing, \\ & HumanBeing(Socrate), \\ & Socrate \text{ teacherOf } Platone \} \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \mathcal{O}' = \{ & Mortal(Socrate), \\ & \mathbf{HumanBeing(Platone)} \\ & \mathbf{Mortal(Platone)} \} \end{aligned}$$

Ontologie nel Web Semantico

Le ontologie definite usando tecnologie del Web Semantico hanno particolari caratteristiche.

Innanzitutto, tutti i nomi sono *URI*:

$$N_C \cup N_P \cup N_I \subseteq \text{URI}.$$

Possono contenere dei *letterali*, che vengono usati per rappresentare tipi di dato *concreti*, come ad esempio stringhe di testo, numeri, date, ...

I letterali possono comparire come *oggetto* di una property assertion. Alcuni esempi di role assertion che coinvolgono letterali sono

Mario surname "Rossi"
Cristiano hasbirth "March – 22 – 1979"

con *Mario*, *Cristiano* $\in N_I$, *surname*, *hasbirth* $\in N_P$ e "Rossi", "March – 22 – 1979" letterali.

È possibile vincolare una proprietà ad avere come codominio solo *letterali* con un certo data type (*datatype property*).

$$\begin{aligned}\text{range}(\textit{surname}) &\sqsubseteq \textit{xsd : string} \\ \text{range}(\textit{hasbirth}) &\sqsubseteq \textit{xsd : date}\end{aligned}$$

Ontologie nel Web Semantico

Le ontologie definite usando tecnologie del Web Semantico hanno particolari caratteristiche.

Innanzitutto, tutti i nomi sono *URI*:

$$N_C \cup N_P \cup N_I \subseteq \text{URI}.$$

Possono contenere dei *letterali*, che vengono usati per rappresentare tipi di dato *concreti*, come ad esempio stringhe di testo, numeri, date, ...

I letterali possono comparire come *oggetto* di una property assertion. Alcuni esempi di role assertion che coinvolgono letterali sono

Mario surname "*Rossi*"
Cristiano hasbirth "*March – 22 – 1979*"

con *Mario*, *Cristiano* $\in N_I$, *surname*, *hasbirth* $\in N_P$ e "*Rossi*", "*March – 22 – 1979*" letterali.

È possibile vincolare una proprietà ad avere come codominio solo *letterali* con un certo data type (*datatype property*).

$\text{range}(\textit{surname}) \sqsubseteq \textit{xsd} : \textit{string}$
 $\text{range}(\textit{hasbirth}) \sqsubseteq \textit{xsd} : \textit{date}$

Ontologie nel Web Semantico

Le ontologie definite usando tecnologie del Web Semantico hanno particolari caratteristiche.

Innanzitutto, tutti i nomi sono *URI*:

$$N_C \cup N_P \cup N_I \subseteq \text{URI}.$$

Possono contenere dei *letterali*, che vengono usati per rappresentare tipi di dato *concreti*, come ad esempio stringhe di testo, numeri, date, ...

I letterali possono comparire come *oggetto* di una property assertion. Alcuni esempi di role assertion che coinvolgono letterali sono

Mario surname "*Rossi*"
Cristiano hasbirth "*March – 22 – 1979*"

con *Mario*, *Cristiano* $\in N_I$, *surname*, *hasbirth* $\in N_P$ e "*Rossi*", "*March – 22 – 1979*" letterali.

È possibile vincolare una proprietà ad avere come codominio solo *letterali* con un certo data type (*datatype property*).

$\text{range}(\textit{surname}) \sqsubseteq \textit{xsd} : \textit{string}$
 $\text{range}(\textit{hasbirth}) \sqsubseteq \textit{xsd} : \textit{date}$

Le ontologie definite usando tecnologie del Web Semantico hanno particolari caratteristiche.

Innanzitutto, tutti i nomi sono *URI*:

$$N_C \cup N_P \cup N_I \subseteq \text{URI}.$$

Possono contenere dei *letterali*, che vengono usati per rappresentare tipi di dato *concreti*, come ad esempio stringhe di testo, numeri, date, ...

I letterali possono comparire come *oggetto* di una property assertion. Alcuni esempi di role assertion che coinvolgono letterali sono

Mario surname "*Rossi*"
Cristiano hasbirth "*March – 22 – 1979*"

con *Mario*, *Cristiano* $\in N_I$, *surname*, *hasbirth* $\in N_P$ e "*Rossi*", "*March – 22 – 1979*" letterali.

È possibile vincolare una proprietà ad avere come codominio solo *letterali* con un certo data type (*datatype property*).

$\text{range}(\textit{surname}) \sqsubseteq \textit{xsd} : \textit{string}$
 $\text{range}(\textit{hasbirth}) \sqsubseteq \textit{xsd} : \textit{date}$

Classi e proprietà vengono raggruppati in *vocabolari* che trattano specifici domini di conoscenza (eg. organizzazioni, pubblica amministrazione, biologia, commercio, etc.).

Un vocabolario può contenere anche alcuni vincoli sulle classi e le proprietà del vocabolario stesso.

Una definizione di vocabolario può essere la seguente:

$$V = (C, P, \Omega)$$

dove

- ① C è un sottoinsieme finito di N_C ,
- ② P è un sottoinsieme finito di N_P ,
- ③ Ω è un insieme finito di vincoli che coinvolgano solo nomi di classi in C e nomi di proprietà in P .

L'utilizzo di vocabolari condivisi (ben noti) favorisce la scalabilità orizzontale delle applicazioni.

Ad esempio, una applicazione sviluppata sull'ontologia di un comune che utilizzi i vocabolari standard per le pubbliche amministrazioni (vedi le *Linee Guida per la Valorizzazione del Patrimonio Informativo Pubblico* dell'*Agenzia per l'Italia Digitale*) può essere estesa senza sforzi aggiuntivi per utilizzare i dati provenienti dalle ontologie di tutti i comuni.

Uno dei primi e più utilizzati vocabolari definiti nell'ambito del Web semantico è *Friend OF A Friend* (FOAF, vedi <http://foaf-project.org>).

FOAF is a project devoted to linking people and information using the Web.

In questa sede ci limiteremo solo alla parte *Core*.

Core - *These classes and properties form the core of FOAF. They describe characteristics of people and social groups that are independent of time and technology; as such they can be used to describe basic information about people in present day, historical, cultural heritage and digital library contexts. In addition to various characteristics of people, FOAF defines classes for Project, Organization and Group as other kinds of agent.*

(tratto da *FOAF Vocabulary Specification 0.99*, Namespace Document 14 January 2014, Paddington Edition, <http://xmlns.com/foaf/spec/>)

Il vocabolario *Foaf Core* è definito come segue:

$\text{FOAFCore} \quad =_{\text{Def}} \quad (C_{\text{foaf}}, P_{\text{foaf}}, \Omega_{\text{foaf}})$

$C_{\text{foaf}} \quad =_{\text{Def}} \quad \{\text{Agent}, \text{Person}, \text{Project}, \text{Organization}, \text{Group}, \text{Document}, \text{Image}\}$

$P_{\text{foaf}} \quad =_{\text{Def}} \quad \{\text{name}, \text{title}, \text{img}, \text{depiction}, \text{depicts}, \text{familyName}, \text{givenName}, \text{based_near}, \text{age}, \text{made}, \text{maker}, \text{primaryTopic}, \text{primaryTopicOf}, \text{member}\}$

$\Omega_{\text{foaf}} \quad =_{\text{Def}} \quad \{\text{Person} \sqsubseteq \text{Agent}, \text{Group} \sqsubseteq \text{Agent}, \text{Organization} \sqsubseteq \text{Agent}, \text{Image} \sqsubseteq \text{Document}, \text{dom}(\text{title}) \sqsubseteq \text{Document}, \text{range}(\text{depiction}) \sqsubseteq \text{Image}, \text{img} \sqsubseteq \text{depiction}, \text{dom}(\text{img}) \sqsubseteq \text{Person}, \text{dom}(\text{knows}) \sqsubseteq \text{Person}, \text{range}(\text{knows}) \sqsubseteq \text{Person}, \dots\}$

Descrizioni Intuitive degli Elementi dei Vocabolari

Le classi e le proprietà di un vocabolario vengono spesso fornite di una descrizione intuitiva nel documento che descrive il vocabolario. Ad esempio, le classi *Agent* e *Person* vengono descritte come segue in <http://xmlns.com/foaf/spec/>

Agent - *The Agent class is the class of agents; things that do stuff. A well known sub-class is Person, representing people. Other kinds of agents include Organization and Group.*

The Agent class is useful in a few places in FOAF where Person would have been overly specific. For example, the IM chat ID properties such as jabberID are typically associated with people, but sometimes belong to software bots.

Person - *The Person class represents people. Something is a Person if it is a person. We don't nitpic about whether they're alive, dead, real, or imaginary. The Person class is a sub-class of the Agent class, since all people are considered 'agents' in FOAF.*

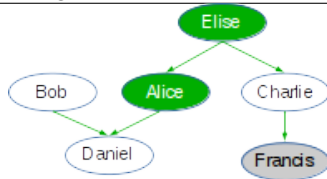
Tuttavia, già nei vincoli di un vocabolario si trovano indicazioni importanti sulla *semantica* dei nomi di classe e di proprietà del vocabolario stesso.

$Person \sqsubseteq Agent$
 $range(depiction) \sqsubseteq Image$
 $img \sqsubseteq depiction,$
 $dom(img) \sqsubseteq Person,$
 $dom(knows) \sqsubseteq Person,$
 $range(knows) \sqsubseteq Person,$
...

Interrogazioni

Il metodo più immediato per ottenere informazioni da una ontologia è il *Conjunctive Query Answering*. Consideriamo ad esempio la seguente ontologia:

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}),$
 $\text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}),$
 $\text{Alice childOf Elise}, \text{Charlie childOf Elise},$
 $\text{Daniel childOf Alice}, \text{Daniel childOf Bob},$
 $\text{Francis childOf Charlie} \}$



Alcune interrogazioni che è possibile effettuare con il conjunctive query answering sono:

- “Trova tutti gli individui maschi.”
- “Chi sono gli individui con almeno un figlio maschio?”
- “Chi sono i figli di *Alice*?”
- “Chi sono gli individui con almeno un figlio maschio ed una femmina?”
- “Chi sono gli individui maschi con almeno un figlio maschio?”

Formule Atomiche

Per definire in maniera rigorosa le query congiuntive è necessario definire preliminarmente l'insieme delle *formule atomiche*.

Sia $V = \{x, y, z, \dots\}$ l'insieme infinito, numerabile e disgiunto da N_C , N_P e N_I delle *variabili*. Le *formule atomiche* sono espressioni dei due seguenti tipi:

$$C(x), \quad P(x, y)$$

con $x, y \in N_I \cup V$, $C \in N_C$ e $P \in N_P$.

Esempi di formule atomiche sono:

- $HumanBeing(x)$,
- $x \text{ childOf Alice}$,
- $Bob \text{ childOf } x$,
- $x \text{ childOf } y$,
- $Mortal(Socrate)$,
- $Alice \text{ childOf Elise}$

con $HumanBeing, Mortal \in N_C$, $childOf \in N_P$, $Alice, Bob, Elise \in N_I$ e $x, y \in V$.

Una formula atomica nella quale non compaiano variabili si dice *chiusa*.

Negli esempi che seguono sono evidenziate le formule atomiche chiuse:

- *HumanBeing*(*x*),
- *x childOf Alice*,
- *Bob childOf x*,
- *x childOf y*,
- *Mortal*(*Socrate*),
- *Alice childOf Elise*

con *HumanBeing*, *Mortal* $\in N_C$, *childOf* $\in N_P$, *Alice*, *Bob*, *Elise* $\in N_I$ e *x*, *y* $\in V$.

Una formula atomica nella quale non compaiano variabili si dice *chiusa*.

Negli esempi che seguono sono evidenziate le formule atomiche chiuse:

- *HumanBeing*(*x*),
- *x childOf Alice*,
- *Bob childOf x*,
- *x childOf y*,
- **Mortal(Socrate)**,
- **Alice childOf Elise**

con *HumanBeing*, *Mortal* $\in N_C$, *childOf* $\in N_P$, *Alice*, *Bob*, *Elise* $\in N_I$ e *x*, *y* $\in V$.

Le asserzioni presenti nelle ontologie sono formule atomiche chiuse.

Una *query congiuntiva* è una congiunzione finita di formule atomiche $T_1 \wedge \dots \wedge T_n$.

Alcuni esempi di query congiuntive:

- “Trova tutti gli individui maschi.”

$$Male(x)$$

- “Chi sono gli individui con almeno un figlio maschio?”

$$y \text{ childOf } x \wedge Male(y)$$

- “Chi sono i figli di *Alice*?”

$$x \text{ childOf } Alice$$

con $x, y \in V$, $Male, Female \in N_C$, $childOf \in N_P$ e $Alice \in N_I$.

Sostituzioni (1/2)

Per definire le *soluzioni* (risposte) delle query congiuntive introduciamo la nozione di *sostituzione*.

Una *sostituzione* $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ ($x_1, \dots, x_n \in V$, $a_1, \dots, a_n \in N_I$) è una mappa finita che associa nomi di individui a variabili.

Sia T una formula atomica e $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione. L'*applicazione* $T\sigma$ di σ a T è la formula atomica che si ottiene sostituendo in T ad ogni occorrenza della variabile x_i il corrispondente nome di individuo a_i , per ogni $1 \leq i \leq n$.

Alcuni esempi:

$$\begin{aligned} \text{Male}(x)[x \rightarrow \text{Bob}] &= \text{Male}(\text{Bob}) \\ \text{Male}(x)[y \rightarrow \text{Bob}] &= \\ (x \text{ childOf } y)[x \rightarrow \text{Alice}] &= \\ (x \text{ childOf } y)[x \rightarrow \text{Alice}, y \rightarrow \text{Elise}] &= \end{aligned}$$

con $x, y \in V$, $\text{Male} \in N_C$, $\text{childOf} \in N_P$ e $\text{Alice}, \text{Bob}, \text{Elise} \in N_I$.

Sostituzioni (1/2)

Per definire le *soluzioni* (risposte) delle query congiuntive introduciamo la nozione di *sostituzione*.

Una *sostituzione* $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ ($x_1, \dots, x_n \in V$, $a_1, \dots, a_n \in N_I$) è una mappa finita che associa nomi di individui a variabili.

Sia T una formula atomica e $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione.
L'*applicazione* $T\sigma$ di σ a T è la formula atomica che si ottiene sostituendo in T ad ogni occorrenza della variabile x_i il corrispondente nome di individuo a_i , per ogni $1 \leq i \leq n$.

Alcuni esempi:

$$\begin{aligned} \text{Male}(x)[x \rightarrow \text{Bob}] &= \text{Male}(\text{Bob}) \\ \text{Male}(x)[y \rightarrow \text{Bob}] &= \text{Male}(x) \\ (x \text{ childOf } y)[x \rightarrow \text{Alice}] &= \\ (x \text{ childOf } y)[x \rightarrow \text{Alice}, y \rightarrow \text{Elise}] &= \end{aligned}$$

con $x, y \in V$, $\text{Male} \in N_C$, $\text{childOf} \in N_P$ e $\text{Alice}, \text{Bob}, \text{Elise} \in N_I$.

Sostituzioni (1/2)

Per definire le *soluzioni* (risposte) delle query congiuntive introduciamo la nozione di *sostituzione*.

Una *sostituzione* $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ ($x_1, \dots, x_n \in V$, $a_1, \dots, a_n \in N_I$) è una mappa finita che associa nomi di individui a variabili.

Sia T una formula atomica e $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione.
L'*applicazione* $T\sigma$ di σ a T è la formula atomica che si ottiene sostituendo in T ad ogni occorrenza della variabile x_i il corrispondente nome di individuo a_i , per ogni $1 \leq i \leq n$.

Alcuni esempi:

$$\begin{aligned} \text{Male}(x)[x \rightarrow \text{Bob}] &= \text{Male}(\text{Bob}) \\ \text{Male}(x)[y \rightarrow \text{Bob}] &= \text{Male}(x) \\ (x \text{ childOf } y)[x \rightarrow \text{Alice}] &= \text{Alice childOf } y \\ (x \text{ childOf } y)[x \rightarrow \text{Alice}, y \rightarrow \text{Elise}] &= \end{aligned}$$

con $x, y \in V$, $\text{Male} \in N_C$, $\text{childOf} \in N_P$ e $\text{Alice}, \text{Bob}, \text{Elise} \in N_I$.

Sostituzioni (1/2)

Per definire le *soluzioni* (risposte) delle query congiuntive introduciamo la nozione di *sostituzione*.

Una *sostituzione* $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ ($x_1, \dots, x_n \in V$, $a_1, \dots, a_n \in N_I$) è una mappa finita che associa nomi di individui a variabili.

Sia T una formula atomica e $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione. L'*applicazione* $T\sigma$ di σ a T è la formula atomica che si ottiene sostituendo in T ad ogni occorrenza della variabile x_i il corrispondente nome di individuo a_i , per ogni $1 \leq i \leq n$.

Alcuni esempi:

$$\begin{array}{ll} \text{Male}(x)[x \rightarrow \text{Bob}] & = \text{Male}(\text{Bob}) \\ \text{Male}(x)[y \rightarrow \text{Bob}] & = \text{Male}(x) \\ (x \text{ childOf } y)[x \rightarrow \text{Alice}] & = \text{Alice childOf } y \\ (x \text{ childOf } y)[x \rightarrow \text{Alice}, y \rightarrow \text{Elise}] & = \text{Alice childOf Elise} \end{array}$$

con $x, y \in V$, $\text{Male} \in N_C$, $\text{childOf} \in N_P$ e $\text{Alice}, \text{Bob}, \text{Elise} \in N_I$.

L'applicazione di sostituzioni a query congiuntive si definisce come segue.

Sia $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione e siano T_1, \dots, T_m formule atomiche. Allora

$$(T_1 \wedge \dots \wedge T_m)\sigma =_{\text{Def}} T_1\sigma \wedge \dots \wedge T_m\sigma.$$

Alcuni esempi:

$$\begin{aligned} (y \text{ childOf } x \wedge \text{Male}(y))[x \rightarrow \text{Alice}] &=_{\text{Def}} \\ (y \text{ childOf } x \wedge \text{Male}(y))[x \rightarrow \text{Alice}, z \rightarrow \text{Bob}] &=_{\text{Def}} \\ (y \text{ childOf } x \wedge \text{Male}(y))[x \rightarrow \text{Alice}, y \rightarrow \text{Bob}] &=_{\text{Def}} \end{aligned}$$

L'applicazione di sostituzioni a query congiuntive si definisce come segue.

Sia $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione e siano T_1, \dots, T_m formule atomiche. Allora

$$(T_1 \wedge \dots \wedge T_m)\sigma =_{\text{Def}} T_1\sigma \wedge \dots \wedge T_m\sigma.$$

Alcuni esempi:

$$\begin{aligned} (y \text{ childOf } x \wedge \text{Male}(y))[x \rightarrow \text{Alice}] &=_{\text{Def}} y \text{ childOf Alice} \wedge \text{Male}(y) \\ (y \text{ childOf } x \wedge \text{Male}(y))[x \rightarrow \text{Alice}, z \rightarrow \text{Bob}] &=_{\text{Def}} \\ (y \text{ childOf } x \wedge \text{Male}(y))[x \rightarrow \text{Alice}, y \rightarrow \text{Bob}] &=_{\text{Def}} \end{aligned}$$

L'applicazione di sostituzioni a query congiuntive si definisce come segue.

Sia $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione e siano T_1, \dots, T_m formule atomiche. Allora

$$(T_1 \wedge \dots \wedge T_m)\sigma =_{\text{Def}} T_1\sigma \wedge \dots \wedge T_m\sigma.$$

Alcuni esempi:

$$\begin{array}{ll} (y \text{ childOf } x \wedge \text{Male}(y))[x \rightarrow \text{Alice}] & =_{\text{Def}} y \text{ childOf Alice} \wedge \text{Male}(y) \\ (y \text{ childOf } x \wedge \text{Male}(y))[x \rightarrow \text{Alice}, z \rightarrow \text{Bob}] & =_{\text{Def}} y \text{ childOf Alice} \wedge \text{Male}(y) \\ (y \text{ childOf } x \wedge \text{Male}(y))[x \rightarrow \text{Alice}, y \rightarrow \text{Bob}] & =_{\text{Def}} \end{array}$$

L'applicazione di sostituzioni a query congiuntive si definisce come segue.

Sia $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione e siano T_1, \dots, T_m formule atomiche. Allora

$$(T_1 \wedge \dots \wedge T_m)\sigma =_{\text{Def}} T_1\sigma \wedge \dots \wedge T_m\sigma.$$

Alcuni esempi:

$(y \text{ childOf } x \wedge \text{Male}(y))[x \rightarrow \text{Alice}]$	$=_{\text{Def}}$	$y \text{ childOf Alice} \wedge \text{Male}(y)$
$(y \text{ childOf } x \wedge \text{Male}(y))[x \rightarrow \text{Alice}, z \rightarrow \text{Bob}]$	$=_{\text{Def}}$	$y \text{ childOf Alice} \wedge \text{Male}(y)$
$(y \text{ childOf } x \wedge \text{Male}(y))[x \rightarrow \text{Alice}, y \rightarrow \text{Bob}]$	$=_{\text{Def}}$	$\text{Bob childOf Alice} \wedge \text{Male}(\text{Bob})$

Soluzioni per una Query

Siano $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione, $Q = T_1 \wedge \dots \wedge T_m$ una query congiuntiva e \mathcal{O} una ontologia.

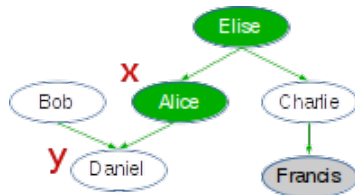
σ è detta essere una *soluzione* per Q rispetto ad \mathcal{O} se e solo se $T_1\sigma, \dots, T_m\sigma$ compaiono in \mathcal{O} .

Consideriamo l'ontologia \mathcal{O} e la query Q definite come segue:

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}), \text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}), \text{Alice childOf Elise}, \text{Charlie childOf Elise}, \text{Daniel childOf Alice}, \text{Daniel childOf Bob}, \text{Francis childOf Charlie} \}$

$Q = y \text{ childOf } x \wedge \text{Male}(y)$

Sia $\sigma_1 = [x \rightarrow \text{Alice}, y \rightarrow \text{Daniel}]$. σ_1 è una soluzione per Q rispetto ad \mathcal{O} ?



Soluzioni per una Query - Esempio 1

Siano $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione, $Q = T_1 \wedge \dots \wedge T_m$ una query congiuntiva e \mathcal{O} una ontologia.

σ è detta essere una *soluzione* per Q rispetto ad \mathcal{O} se e solo se $T_1\sigma, \dots, T_m\sigma$ compaiono in \mathcal{O} .

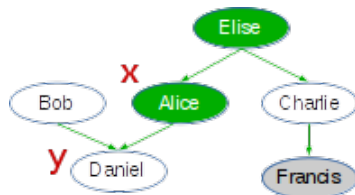
Consideriamo l'ontologia \mathcal{O} e la query Q ("Chi sono gli individui con almeno un figlio maschio?") definite come segue:

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}), \text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}), \text{Alice childOf Elise}, \text{Charlie childOf Elise}, \text{Daniel childOf Alice}, \text{Daniel childOf Bob}, \text{Francis childOf Charlie} \}$

$Q = y \text{ childOf } x \wedge \text{Male}(y)$

Sia $\sigma_1 = [x \rightarrow \text{Alice}, y \rightarrow \text{Daniel}]$. σ_1 è una soluzione per Q rispetto ad \mathcal{O} ? **SI**.

$Q\sigma_1 = \text{Daniel childOf Alice} \wedge \text{Male}(\text{Daniel})$.



Soluzioni per una Query - Esempio 2

Siano $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione, $Q = T_1 \wedge \dots \wedge T_m$ una query congiuntiva e \mathcal{O} una ontologia.

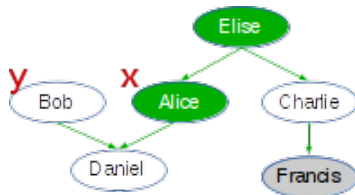
σ è detta essere una *soluzione* per Q rispetto ad \mathcal{O} se e solo se $T_1\sigma, \dots, T_m\sigma$ compaiono in \mathcal{O} .

Consideriamo l'ontologia \mathcal{O} e la query Q definite come segue:

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}), \text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}), \text{Alice childOf Elise}, \text{Charlie childOf Elise}, \text{Daniel childOf Alice}, \text{Daniel childOf Bob}, \text{Francis childOf Charlie} \}$

$Q = y \text{ childOf } x \wedge \text{Male}(y).$

Sia $\sigma_2 = [x \rightarrow \text{Alice}, y \rightarrow \text{Bob}]$. σ_2 è una soluzione per Q rispetto ad \mathcal{O} ?



Soluzioni per una Query - Esempio 2

Siano $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione, $Q = T_1 \wedge \dots \wedge T_m$ una query congiuntiva e \mathcal{O} una ontologia.

σ è detta essere una *soluzione* per Q rispetto ad \mathcal{O} se e solo se $T_1\sigma, \dots, T_m\sigma$ compaiono in \mathcal{O} .

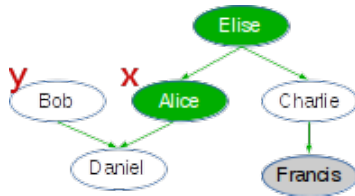
Consideriamo l'ontologia \mathcal{O} e la query Q ("Chi sono gli individui con almeno un figlio maschio?") definite come segue:

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}), \text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}), \text{Alice childOf Elise}, \text{Charlie childOf Elise}, \text{Daniel childOf Alice}, \text{Daniel childOf Bob}, \text{Francis childOf Charlie} \}$

$Q = y \text{ childOf } x \wedge \text{Male}(y).$

Sia $\sigma_2 = [x \rightarrow \text{Alice}, y \rightarrow \text{Bob}]$. σ_2 è una soluzione per Q rispetto ad \mathcal{O} ? **NO**.

$Q\sigma_2 = \text{Bob childOf Alice} \wedge \text{Male}(\text{Bob}).$



Soluzioni per una Query - Esempio 3

Siano $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione, $Q = T_1 \wedge \dots \wedge T_m$ una query congiuntiva e \mathcal{O} una ontologia.

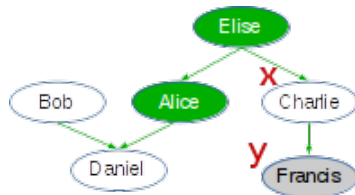
σ è detta essere una *soluzione* per Q rispetto ad \mathcal{O} se e solo se $T_1\sigma, \dots, T_m\sigma$ compaiono in \mathcal{O} .

Consideriamo l'ontologia \mathcal{O} e la query Q definite come segue:

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}), \text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}), \text{Alice childOf Elise}, \text{Charlie childOf Elise}, \text{Daniel childOf Alice}, \text{Daniel childOf Bob}, \text{Francis childOf Charlie} \}$

$Q = y \text{ childOf } x \wedge \text{Male}(y).$

Sia $\sigma_3 = [x \rightarrow \text{Charlie}, y \rightarrow \text{Francis}]$. σ_3 è una soluzione per Q rispetto ad \mathcal{O} ?



Soluzioni per una Query - Esempio 3

Siano $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione, $Q = T_1 \wedge \dots \wedge T_m$ una query congiuntiva e \mathcal{O} una ontologia.

σ è detta essere una *soluzione* per Q rispetto ad \mathcal{O} se e solo se $T_1\sigma, \dots, T_m\sigma$ compaiono in \mathcal{O} .

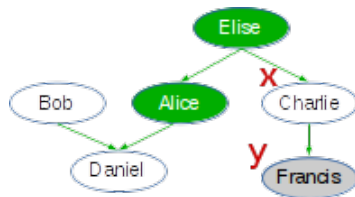
Consideriamo l'ontologia \mathcal{O} e la query Q ("Chi sono gli individui con almeno un figlio maschio?") definite come segue:

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}), \text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}), \text{Alice childOf Elise}, \text{Charlie childOf Elise}, \text{Daniel childOf Alice}, \text{Daniel childOf Bob}, \text{Francis childOf Charlie} \}$

$Q = y \text{ childOf } x \wedge \text{Male}(y).$

Sia $\sigma_3 = [x \rightarrow \text{Charlie}, y \rightarrow \text{Francis}]$. σ_2 è una soluzione per Q rispetto ad \mathcal{O} ? **NO**.

$Q\sigma_3 = \text{Francis childOf Charlie} \wedge \text{Male}(\text{Francis}).$



Soluzioni Minimali per una Query

Siano $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione, Q una query congiuntiva e \mathcal{O} una ontologia.

σ è una *soluzione minimale* per Q rispetto a \mathcal{O} se e solo se:

- 1 σ è una soluzione per Q rispetto ad \mathcal{O} e inoltre
- 2 tutte le variabili x_1, \dots, x_n che compaiono in σ compaiono anche in Q (criterio di minimalità).

Consideriamo ad esempio

$$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}), \\ \text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}), \\ \text{Alice childOf Elise}, \text{Charlie childOf Elise}, \\ \text{Daniel childOf Alice}, \text{Daniel childOf Bob}, \\ \text{Francis childOf Charlie} \}$$
$$Q = y \text{ childOf } x \wedge \text{Male}(y).$$

$\sigma_5 = [x \rightarrow \text{Elise}, y \rightarrow \text{Charlie}, z \rightarrow \text{Francis}]$ è una soluzione minimale per Q rispetto a \mathcal{O} ?

Soluzioni Minimali per una Query

Siano $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione, Q una query congiuntiva e \mathcal{O} una ontologia.

σ è una *soluzione minimale* per Q rispetto a \mathcal{O} se e solo se:

- 1 σ è una soluzione per Q rispetto ad \mathcal{O} e inoltre
- 2 tutte le variabili x_1, \dots, x_n che compaiono in σ compaiono anche in Q (criterio di minimalità).

Consideriamo ad esempio

$$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}), \\ \text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}), \\ \text{Alice childOf Elise}, \text{Charlie childOf Elise}, \\ \text{Daniel childOf Alice}, \text{Daniel childOf Bob}, \\ \text{Francis childOf Charlie} \}$$
$$Q = y \text{ childOf } x \wedge \text{Male}(y).$$

$\sigma_5 = [x \rightarrow \text{Elise}, y \rightarrow \text{Charlie}, z \rightarrow \text{Francis}]$ è una soluzione minimale per Q rispetto a \mathcal{O} ? **NO**.

$\sigma_6 = [x \rightarrow \text{Elise}, y \rightarrow \text{Charlie}]$ è una soluzione minimale per Q rispetto a \mathcal{O} ? .

Soluzioni Minimali per una Query

Siano $\sigma = [x_1 \rightarrow a_1, \dots, x_n \rightarrow a_n]$ una sostituzione, Q una query congiuntiva e \mathcal{O} una ontologia.

σ è una *soluzione minimale* per Q rispetto a \mathcal{O} se e solo se:

- 1 σ è una soluzione per Q rispetto ad \mathcal{O} e inoltre
- 2 tutte le variabili x_1, \dots, x_n che compaiono in σ compaiono anche in Q (criterio di minimalità).

Consideriamo ad esempio

$$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}), \\ \text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}), \\ \text{Alice childOf Elise}, \text{Charlie childOf Elise}, \\ \text{Daniel childOf Alice}, \text{Daniel childOf Bob}, \\ \text{Francis childOf Charlie} \}$$
$$Q = y \text{ childOf } x \wedge \text{Male}(y).$$

$\sigma_5 = [x \rightarrow \text{Elise}, y \rightarrow \text{Charlie}, z \rightarrow \text{Francis}]$ è una soluzione minimale per Q rispetto a \mathcal{O} ? **NO**.

$\sigma_6 = [x \rightarrow \text{Elise}, y \rightarrow \text{Charlie}]$ è una soluzione minimale per Q rispetto a \mathcal{O} ? **SI**.

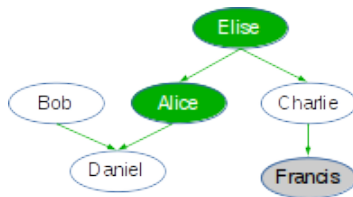
Conjunctive Query Answering

Il problema del *Conjunctive Query Answering* consiste nel trovare tutte le soluzioni minimali di una query congiuntiva rispetto ad una ontologia.

Esempio 1: "Trova tutti gli individui maschi."

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}), \text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}), \text{Alice childOf Elise}, \text{Charlie childOf Elise}, \text{Daniel childOf Alice}, \text{Daniel childOf Bob}, \text{Francis childOf Charlie} \}$

$Q = \text{Male}(x)$



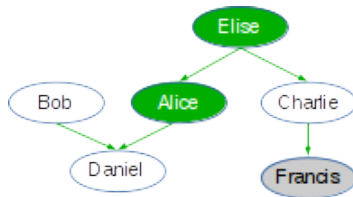
x
Bob
Charlie
Daniel

Conjunctive Query Answering - Esempio 2

“Chi sono gli individui con almeno un figlio maschio?”

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}),$
 $\text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}),$
 $\text{Alice childOf Elise}, \text{Charlie childOf Elise},$
 $\text{Daniel childOf Alice}, \text{Daniel childOf Bob},$
 $\text{Francis childOf Charlie} \}$

$Q = y \text{ childOf } x \wedge \text{Male}(y)$



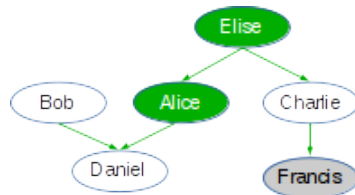
x	y
Elise	Charlie
Alice	Daniel
Bob	Daniel

Conjunctive Query Answering - Esempio 3

“Chi sono i figli di *Elise*?”

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}),$
 $\text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}),$
 $\text{Alice childOf Elise}, \text{Charlie childOf Elise},$
 $\text{Daniel childOf Alice}, \text{Daniel childOf Bob},$
 $\text{Francis childOf Charlie} \}$

$Q = x \text{ childOf Elise}$



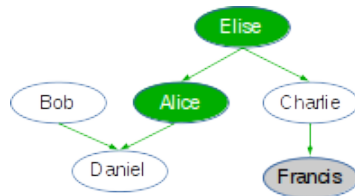
x

Conjunctive Query Answering - Esempio 3

“Chi sono i figli di *Elise*?”

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}),$
 $\text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}),$
 $\text{Alice childOf Elise}, \text{Charlie childOf Elise},$
 $\text{Daniel childOf Alice}, \text{Daniel childOf Bob},$
 $\text{Francis childOf Charlie} \}$

$Q = x \text{ childOf Elise}$

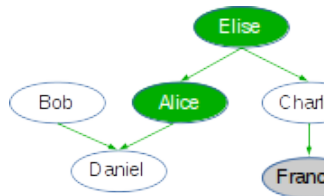


x
Alice
Charlie

Conjunctive Query Answering - Esempio 4

“Chi sono gli individui con almeno un figlio maschio ed una femmina?”

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}),$
 $\text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}),$
 $\text{Alice childOf Elise}, \text{Charlie childOf Elise},$
 $\text{Daniel childOf Alice}, \text{Daniel childOf Bob},$
 $\text{Francis childOf Charlie} \}$



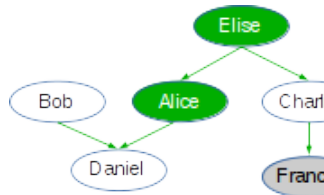
x	y	z

Conjunctive Query Answering - Esempio 4

“Chi sono gli individui con almeno un figlio maschio ed una femmina?”

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}),$
 $\text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}),$
 $\text{Alice childOf Elise}, \text{Charlie childOf Elise},$
 $\text{Daniel childOf Alice}, \text{Daniel childOf Bob},$
 $\text{Francis childOf Charlie} \}$

$Q = y \text{ childOf } x \wedge z \text{ childOf } x \wedge \text{Male}(y) \wedge \text{Female}(z)$



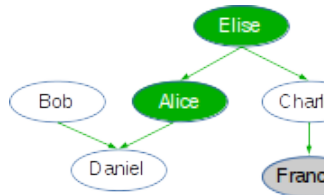
x	y	z

Conjunctive Query Answering - Esempio 4

“Chi sono gli individui con almeno un figlio maschio ed una femmina?”

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}),$
 $\text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}),$
 $\text{Alice childOf Elise}, \text{Charlie childOf Elise},$
 $\text{Daniel childOf Alice}, \text{Daniel childOf Bob},$
 $\text{Francis childOf Charlie} \}$

$Q = y \text{ childOf } x \wedge z \text{ childOf } x \wedge \text{Male}(y) \wedge \text{Female}(z)$

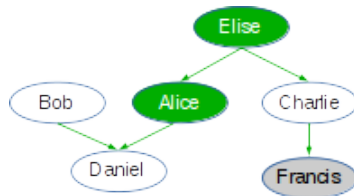


x	y	z
Elise	Charlie	Alice

Conjunctive Query Answering - Esempio 5

“Chi sono gli individui maschi con almeno una figlia femmina?”

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}),$
 $\text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}),$
 $\text{Alice childOf Elise}, \text{Charlie childOf Elise},$
 $\text{Daniel childOf Alice}, \text{Daniel childOf Bob},$
 $\text{Francis childOf Charlie} \}$



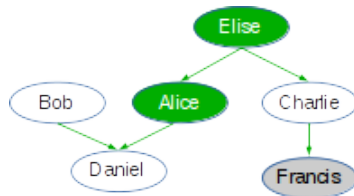
x	y

Conjunctive Query Answering - Esempio 5

“Chi sono gli individui maschi con almeno una figlia femmina?”

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}),$
 $\text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}),$
 $\text{Alice childOf Elise}, \text{Charlie childOf Elise},$
 $\text{Daniel childOf Alice}, \text{Daniel childOf Bob},$
 $\text{Francis childOf Charlie} \}$

$Q = \text{Male}(x) \wedge y \text{ childOf } x \wedge \text{Female}(y)$



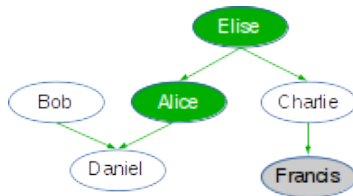
x	y

Conjunctive Query Answering - Esempio 5

“Chi sono gli individui maschi con almeno una figlia femmina?”

$\mathcal{O} = \{ \text{Female}(\text{Elise}), \text{Female}(\text{Alice}), \text{Male}(\text{Bob}), \text{Male}(\text{Charlie}), \text{Male}(\text{Daniel}), \text{Alice childOf Elise}, \text{Charlie childOf Elise}, \text{Daniel childOf Alice}, \text{Daniel childOf Bob}, \text{Francis childOf Charlie} \}$

$Q = \text{Male}(x) \wedge y \text{ childOf } x \wedge \text{Female}(y)$



Nessuna soluzione

Il protocollo SPARQL

Le basi di conoscenza presenti sul Web Semantico usualmente mettono a disposizione uno *SPARQL endpoint* che permette di interrogarle e, ove permesso, di modificarle.

Knowledge Base	Endpoint IRI
Europeana	http://europeana.ontotext.com/sparql
CNR	http://data.cnr.it/sparql/
Camera dei Deputati	http://dati.camera.it/sparql
DBPedia	http://dbpedia.org/sparql

Table : Alcuni endpoint sparql

Il *protocollo SPARQL* (vedi <http://www.w3.org/TR/sparql11-protocol/>) è basato sul protocollo HTTP le richieste SPARQL vengono inviate agli endpoint come richieste GET o POST e l'endpoint risponde con un *esito*.

Il protocollo SPARQL

Le basi di conoscenza presenti sul Web Semantico usualmente mettono a disposizione uno *SPARQL endpoint* che permette di interrogarle e, ove permesso, di modificarle.

Knowledge Base	Endpoint IRI
Europeana	http://europeana.ontotext.com/sparql
CNR	http://data.cnr.it/sparql/
Camera dei Deputati	http://dati.camera.it/sparql
DBPedia	http://dbpedia.org/sparql

Table : Alcuni endpoint sparql

Il *protocollo SPARQL* (vedi <http://www.w3.org/TR/sparql11-protocol/>) è basato sul protocollo HTTP le richieste SPARQL vengono inviate agli endpoint come richieste GET o POST e l'endpoint risponde con un *esito*.

Le richieste di tipo *query* vanno specificate nel linguaggio denominato *SPARQL Query*. La specifica di questo linguaggio è disponibile all'indirizzo

<http://www.w3.org/TR/sparql11-query/> .

Una *query SPARQL* ha la seguente sintassi

```
SELECT ?x1 ... ?xm WHERE { Q }
```

dove:

- $?x_1, \dots, ?x_m$ sono *variabili* ($m > 0$);
- Q è una query congiuntiva nella quale compaiono tutte le variabili $?x_1, \dots, ?x_m$;
- le formule atomiche di tipo $C(?x)$ sono scritte come $?x$ a C .

Alcuni esempi di query SPARQL sono:

- “Trova tutti gli individui maschi.”

```
SELECT ?x WHERE { ?x a <http://example.org/Male> }
```

- “Chi sono gli individui con almeno un figlio maschio?”

```
SELECT ?x WHERE { ?y <http://example.org/childOf> ?x .  
                  ?y a <http://example.org/Male> }
```

- “Chi sono i figli di *Alice*?”

```
SELECT ?x WHERE { ?x <http://example.org/childOf> <http://example.org/Alice> }
```