Web semantico e logiche descrittive

M. Simi, 2011-2012
Cap 2 del "Description Logic Handbook"
Lezioni di U.Straccia

Categorie e oggetti

- Molti dei ragionamenti che si fanno sono sulle categorie piuttosto che sugli individui
- Se organizziamo la conoscenza in categorie (e sottocategorie) è sufficiente classificare un oggetto, tramite le proprietà percepite, per inferire le proprietà della categoria|e a cui appartiene (ereditarietà)

Ontologie di dominio

- Molte delle idee delle reti semantiche e dei frame sono state raccolte in logiche specializzate
- Queste logiche sono alla base delle proposte per il *Web semantico*
- *Ontologia*: modello formale di un dominio di interesse (una concettualizzazione)
- Le relazioni di sottoclasse organizzano la conoscenza in tassonomie (come in botanica, biologia, nelle scienze librarie …)

Il Web semantico

- La visione di Tim Berners-Lee (1998): da un Web "sintattico" per la comunicazione tra persone al Web "semantico", una grossa rete di informazioni collegate su scala globale comprensibili ai programmi (un database globale – *linked data*)
- Il veicolo è XML, un linguaggio di annotazione generico.

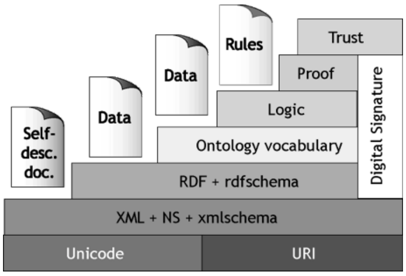
Le motivazioni del web semantico

- Il web di oggi, fatto di documenti, è adatto alla fruizione da parte delle persone
  - i motori di ricerca sono sintattici e i risultati non sono utilizzabili dai programmi
  - Impossibile|difficile costruire dei programmi che fanno uso efficace dei contenuti web
- Il web semantico si propone di aiutare nella gestione della conoscenza presente nel web
- Un **web di dati** da condividere e riutilizzare tra applicazioni, aziende e comunità

Web semantico e gestione della conoscenza

- Ricerca dell'informazione
  - Ricerche meno sintattiche
- Estrazione dell'informazione
  - Adesso sono le persone che ricercano, interpretano e combinano i contenuti delle pagine web
- Manutenzione dell'informazione
  - Terminologie inconsistenti, informazioni obsolete
- Presentazione dell'informazione
  - Impossibile definere "viste" sui contenuti web

Il livelli del Web semantico



I livelli del web semantico

- Unicode e URI
- XML interoperabilità sintattica
- RDF (Resource Description Framework): per descrivere relazioni semantiche tra risorse (*soggetto, predicato, oggetto*)
- RDF schema (RDFS): per vincolare domini e codomini delle relazioni, definire classi di oggetti, relazioni tra classi; RDFS linguaggio per ontologie, poco espressivo

Web semantico e linguaggi per ontologie

- Linguaggi per l'aggiunta di un servizio inferenziale a RDF:
  - OIL gruppo europeo
  - DAML-ONT gruppo americano
  - DAML+OIL proposto come standard
  - OWL: Web Ontology Language, standard W3C.
- OWL evolve dalle logiche descrittive

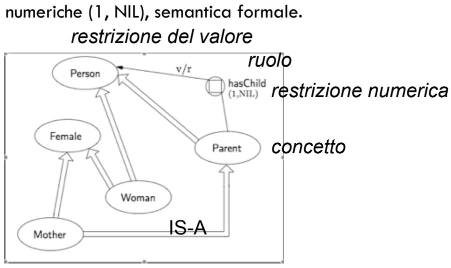
Logiche descrittive

Possono essere viste come:

- *Evoluzioni "logiche"* di linguaggi di KR "a rete", come i *frame* e le *reti semantiche*
- *Contrazioni* della logica del prim'ordine (FOL) per ottenere migliori proprietà computazionali

KL-One

- KL-One [Brachman-Schmolze 1985] introduce le idee importanti delle DL: concetti e ruoli, restrizioni sui valori, restrizioni numeriche (1, NIL), semantica formale.



DL come formalizzazione di reti semantiche

- Verso gli anni 80 si ha una sterzata verso la logica delle reti semantiche
- Il processo consiste nel
  - riformulare i costrutti secondo i canoni della logica
  - eliminare i costrutti che non si prestano a tale riformulazione (*default* ed eccezioni)

Da KL-One alle logiche descrittive

- Logiche terminologiche
  - $\mathcal{FL}$- (Frame Language) [Brachman and Levesque,1984] tradeoff tra espressività di un linguaggio di rappresentazione e la complessità del ragionamento
  - CLASSIC [Brachman 1991], limitato e completo
- Logiche descrittive
  - LOOM [MacGregor-Bates 1987], BACK [Nebel-vonLuck, 1988], espressivi e incompleti
  - KRIS [Baader, Hollunder, 1991], espressivi e completi
  - FaCT, DLP, Racer, sistemi ottimizzati per logiche espressive

Esempio

La seguente è una formula di una delle DL:

(and Paper (atmost 2 hasAuthor)
    (atleast 2 hasAuthor)) [paper3]

paper3: Paper $\sqcap$ ($\leq$ 2 hasAuthor) $\sqcap$ ($\geq$ 2 hasAuthor)

(sintassi alternativa)

equivalente a:

Paper(paper3) $\wedge$

$\exists x$ hasAuthor(paper3, x) $\wedge$

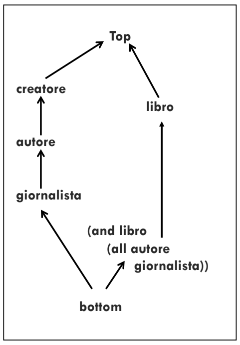
$\exists y$ hasAuthor(paper3, y) $\wedge$ $x \neq y$ $\wedge$

hasAuthor(paper3, z) $\Rightarrow$ ($z = x$) $\vee$ ($z = y$)

Concetti, ruoli, individui

- Ogni DL è caratterizzata da operatori per la costruzione di termini di due tipi:
  - *Concetti,* corrispondenti a relazioni unarie
    con operatori per la costruzione di concetti complessi: and, or , not, all, some, atleast, atmost, …
  - *Ruoli,* corrispondenti a relazioni binarie
    ed eventualmente operatori
- Individui: usati solo nelle asserzioni

Una KB basata su logica descrittiva



KB

T-BOX

giornalista $\Rightarrow$ autore
articolo $\Leftrightarrow$
(**and** (**a-not** libro)
(**all** autore giornalista))
autore $\Rightarrow$ creatore

A-BOX

autore[Eco, I1]
autore[Biagi, I2]
giornalista[Biagi]
(**and** libro
(**all** autore
giornalista))[a2]

La logica $\mathcal{AL}$: la sintassi dei termini

<concetto> $\rightarrow$	A		
	| $\top$	(top, concetto universale)	
	| $\bot$	(bottom)	
	| $\neg A$	(negazione atomica)	
	| $C \sqcap D$	(intersezione)	
	| $\forall R.C$	(restrizione di valore)	
	| $\exists R.\top$	(esistenziale debole)	

<ruolo> $\rightarrow R$

A, B concetti primitivi   R ruolo primitivo

C, D concetti

Esempi

- Person $\sqcap$ Female
- Person $\sqcap$ $\neg$Female
- Person $\sqcap$ $\exists$ hasChild.T
- Person $\sqcap$ $\forall$ hasChild.Female
- Person $\sqcap$ $\forall$ hasChild.$\bot$

Semantica di $\mathcal{AL}$

$\Delta^{\mathcal{I}}$ dominio di interpretazione, un insieme di individui

$\mathcal{I}$ funzione di interpretazione che assegna a:

- concetti atomici *A*: $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
- ruoli atomici *R*: $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
- nomi di individuo *a*: $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$

$\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$

$\bot^{\mathcal{I}} = \emptyset$

$(\neg A)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus A^{\mathcal{I}}$ *il complemento*

$(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$

$(\forall R.C)^{\mathcal{I}} = \{a \in \Delta^{\mathcal{I}} \mid \forall b.(a, b) \in R^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}}\}$

$(\exists R.\top)^{\mathcal{I}} = \{a \in \Delta^{\mathcal{I}} \mid \exists b.(a, b) \in R^{\mathcal{I}}\}$

Definizioni

- Un'interpretazione $\mathcal{I}$ è un *modello* di un concetto C sse $C^{\mathcal{I}} \neq \emptyset$ (insieme vuoto). Lo stesso per i ruoli.
- C ≡ *D (equivalenti) sse* $C^{\mathcal{I}} = D^{\mathcal{I}}$ per ogni $\mathcal{I}$
- Esempio:

  ∀ haFiglio.Femmina ⊓ ∀ haFiglio.Studente ≡

  ∀ haFiglio. Femmina ⊓ Studente

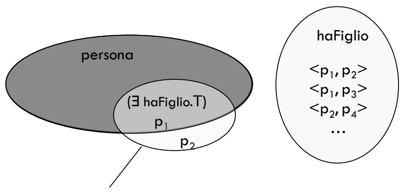
Esempio 1

Articolo ⊓ ∃ haAutore.⊤ ⊓
∀ haAutore.Giornalista

"*l'insieme degli articoli che hanno almeno un autore, e i cui autori sono tutti giornalisti*"

Esempio 2

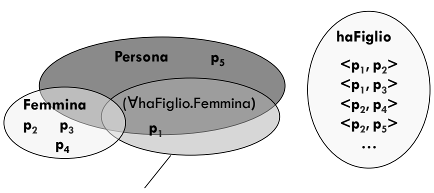
Persona ⊓ ∃haFiglio.⊤



$\{ a \in \Delta^{I} \mid \exists b\ (a, b) \in \text{haFiglio}^{I}\}$

Esempio 3

Persona ⊓ ∀haFiglio.Femmina



$\{a \in \Delta^{I} \mid \forall b\ (a, b) \in \text{haFiglio}) \Rightarrow b \in \text{Femmina}^{I}\}$

Logiche più espressive

$\mathcal{U}$: unione, $(C \sqcup D)^{\mathcal{I}} = (C^{\mathcal{I}} \cup D^{\mathcal{I}})$

$\mathcal{E}$: esistenziale pieno

$(\exists R.C)^{\mathcal{I}} = \{a \in \Delta^{I} \mid \exists b.(a, b) \in R^{I} \wedge b \in C^{I}\}$

$\mathcal{N}$: restrizioni numeriche

$(\geq n\ R)^{\mathcal{I}} = \{a \in \Delta^{I} \mid |\{ b \mid (a, b) \in R^{I}\}| \geq n\}$ (atleast)

$(\leq n\ R)^{\mathcal{I}} = \{a \in \Delta^{I} \mid |\{ b \mid (a, b) \in R^{I}\}| \leq n\}$ (atmost)

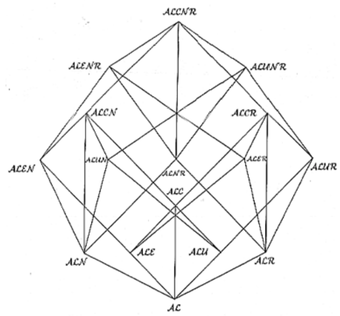
n, numero intero | . | cardinalità dell'insieme

$\mathcal{C}$: complemento pieno, $(\neg C)^{I} = \Delta^{I} \setminus C^{I}$

La famiglia $\mathcal{AL}$

- Le diverse DL sono ottenute aggiungendo altri costruttori di termini ad $\mathcal{AL}$
  $\mathcal{AL}[\mathcal{U}][\mathcal{E}][\mathcal{N}][\mathcal{C}]$
- Non tutte distinte
  - $\mathcal{ALUE} = \mathcal{ALC}$ dato che $(C \sqcup D) \equiv \neg(\neg C \sqcap \neg D)$ e $\exists R.C \equiv \neg\forall R.\neg C$
  - $\mathcal{ALCN} = \mathcal{ALUEN}$

Il reticolo della famiglia $\mathcal{AL}$



Il linguaggio della T-BOX

- Assiomi terminologici $\mathcal{T}$

$C \sqsubseteq D$	*inclusione di concetti*, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
$R \sqsubseteq S$	*inclusione di ruoli*, $R^{\mathcal{I}} \subseteq S^{\mathcal{I}}$
$C \equiv D$	*uguaglianza di concetti*, $C^{\mathcal{I}} \equiv D^{\mathcal{I}}$
$R \equiv S$	*uguaglianza di ruoli*, $R^{\mathcal{I}} \equiv S^{\mathcal{I}}$

$\mathcal{I}$ soddisfa $\mathcal{T}$ sse soddisfa ogni elemento in $\mathcal{T}$

Terminologia (T-BOX)

- *Definizioni*: uguaglianze che introducono un simbolo sulla sinistra

  $Mother \equiv Woman \sqcap hasChild.Person$
- *Terminologia*: i simboli compaiono sulla sinistra non più di una volta.
- *Simboli primitivi*: compaiono solo a destra
- Simboli *definiti*: compaiono anche sulla sinistra
- Assumiamo $\mathcal{T}$ *acicliche*.

Una terminologia aciclica

Woman	$\equiv$	Person $\sqcap$ Female
Man	$\equiv$	Person $\sqcap \neg$ $\boxed{\text{Woman}}$
Mother	$\equiv$	Woman $\sqcap\ \exists$hasChild.Person
Father	$\equiv$	$\boxed{\text{Man}}$ $\sqcap\ \exists$hasChild.Person
Parent	$\equiv$	$\boxed{\text{Father}}$ $\sqcup$ $\boxed{\text{Mother}}$
Grandmother	$\equiv$	$\boxed{\text{Mother}}$ $\sqcap\ \exists$hasChild.$\boxed{\text{Parent}}$
MotherWithManyChildren	$\equiv$	$\boxed{\text{Mother}}$ $\sqcap \geqslant 3$ hasChild
MotherWithoutDaughter	$\equiv$	$\boxed{\text{Mother}}$ $\sqcap\ \forall$hasChild.$\neg$ $\boxed{\text{Woman}}$
Wife	$\equiv$	$\boxed{\text{Woman}}$ $\sqcap\ \exists$hasHusband.$\boxed{\text{Man}}$

Espansione di $\mathcal{T}$

- Se una terminologia è *aciclica* può essere *espansa* sostituendo ai simboli definiti le loro definizioni.
- Il processo converge e l'espansione, $\mathcal{T}^e$, è unica.
- Proprietà di $\mathcal{T}^e$:
  - in $\mathcal{T}^e$ ogni uguaglianza ha la forma $C \equiv D^e$ *dove* D^e contiene solo simboli primitivi
  - $\mathcal{T}^e$ contiene gli stessi simboli primitivi e definiti di $\mathcal{T}$
  - $\mathcal{T}^e$ è equivalente a $\mathcal{T}$

Esempio: espansione

$$
\begin{aligned}
\text{Woman} &\equiv \text{Person} \sqcap \text{Female} \\
\text{Man} &\equiv \text{Person} \sqcap \neg(\text{Person} \sqcap \text{Female}) \\
\text{Mother} &\equiv (\text{Person} \sqcap \text{Female}) \sqcap \exists \text{hasChild.Person} \\
\text{Father} &\equiv (\text{Person} \sqcap \neg(\text{Person} \sqcap \text{Female})) \sqcap \exists \text{hasChild.Person} \\
\text{Parent} &\equiv ((\text{Person} \sqcap \neg(\text{Person} \sqcap \text{Female})) \sqcap \exists \text{hasChild.Person}) \\
&\quad \sqcup ((\text{Person} \sqcap \text{Female}) \sqcap \exists \text{hasChild.Person}) \\
\text{Grandmother} &\equiv ((\text{Person} \sqcap \text{Female}) \sqcap \exists \text{hasChild.Person}) \\
&\quad \sqcap \exists \text{hasChild.}(((\text{Person} \sqcap \neg(\text{Person} \sqcap \text{Female})) \\
&\qquad \sqcap \exists \text{hasChild.Person}) \\
&\qquad \sqcup ((\text{Person} \sqcap \text{Female}) \\
&\qquad \sqcap \exists \text{hasChild.Person})) \\
\text{MotherWithManyChildren} &\equiv ((\text{Person} \sqcap \text{Female}) \sqcap \exists \text{hasChild.Person}) \sqcap \geqslant 3\ \text{hasChild} \\
\text{MotherWithoutDaughter} &\equiv ((\text{Person} \sqcap \text{Female}) \sqcap \exists \text{hasChild.Person}) \\
&\quad \sqcap \forall \text{hasChild.}(\neg(\text{Person} \sqcap \text{Female})) \\
\text{Wife} &\equiv (\text{Person} \sqcap \text{Female}) \\
&\quad \sqcap \exists \text{hasHusband.}(\text{Person} \sqcap \neg(\text{Person} \sqcap \text{Female}))
\end{aligned}
$$

Assiomi di inclusione

- *Specializzazione*: un assioma di inclusione la cui parte sinistra è atomica.
  Woman $\sqsubseteq$ Person
- *Normalizzazione*: Una terminologia *generalizzata* [con assiomi di inclusione], se aciclica, può essere trasformata in una terminologia equivalente con solo definizioni:
  $A \sqsubseteq C \rightarrow A \equiv A' \sqcap C$

con A' nuovo simbolo primitivo

Proprietà della normalizzazione

- Sia T una terminologia generalizzata e T' una sua normalizzazione.

1. Ogni modello di T' è un modello di T.
2. Per ogni modello $\mathcal{I}$ di T c'è un modello $\mathcal{I}'$ di T' che ha lo stesso dominio di $\mathcal{I}$ e coincide con $\mathcal{I}$ sui concetti e ruoli atomici in T

- Le inclusioni non aggiungono potere espressivo nel caso di terminologie acicliche

Il linguaggio delle asserzioni: A-BOX

- Una A-BOX è un insieme di asserzioni di due tipi:
  a:C, asserzioni su concetti, $a^{\mathcal{I}} \in C^{\mathcal{I}}$
  (b, c):R, asserzioni su ruoli, $(b^{\mathcal{I}}, c^{\mathcal{I}}) \in R^{\mathcal{I}}$
  - a, b, c, d … sono meta-simboli per individui
  - $\mathcal{I}$ fornisce anche una interpretazione per i simboli di individuo

Esempio di A-BOX

Mary: Mother	Peter: Father
(Mary, Peter): hasChild	(Peter, Harry): hasChild
(Mary, Paul): hasChild	

- Assunzione di *mondo aperto* (OWA): non si assume di specificare tutto
- Assunzione di *nome unico* (UNA): simboli diversi, individui diversi

DL come frammenti del FOL

- È possibile convertire asserzioni delle logiche descrittive in formule FOL
- La traduzione avviene attraverso la definizione di una funzione di traduzione t (C, x) che restituisce una formula del FOL con x libera
  t (C, x) $\mapsto$ C(x)

Traduzione da DL a FOL

$t\,(C \sqsubseteq D)$	$\mapsto$	$\forall x\,.\; t\,(C, x) \Rightarrow t\,(D, x)$
$t\,(a{:}C)$	$\mapsto$	$t\,(C, a)$
$t\,((a, b){:}R)$	$\mapsto$	$R(a, b)$
$t\,(\top, x)$	$\mapsto$	true
$t\,(\bot, x)$	$\mapsto$	false
$t\,(A, x)$	$\mapsto$	$A(x)$
$t\,(C \sqcap D, x)$	$\mapsto$	$t\,(C, x) \wedge t\,(D, x)$
$t\,(C \sqcup D, x)$	$\mapsto$	$t\,(C, x) \vee t\,(D, x)$

Traduzione da DL a FOL (cont.)

$t\,(\neg C, x)$	$\mapsto$	$\neg t\,(C, x)$
$t\,(\exists R.C, x)$	$\mapsto$	$\exists y\,.\; R(x, y) \wedge t\,(C, y)$
$t\,(\forall R.C, x)$	$\mapsto$	$\forall y\,.\; R(x, y) \Rightarrow t\,(C, y)$

Esempi di traduzione

$t\,(\text{HappyFather} \sqsubseteq \text{Man} \sqcap \exists\,\text{hasChild}\,.\,\text{Female}) =$
$\forall x\,.\; t\,(\text{HappyFather}, x) \Rightarrow t\,(\text{Man} \sqcap \exists\,\text{hasChild}\,.\,\text{Female}, x) =$
$\forall x\,.\; \text{HappyFather}(x) \Rightarrow t\,(\text{Man}, x) \wedge t\,(\exists\,\text{hasChild}\,.\,\text{Female}, x) =$
$\forall x\,.\; \text{HappyFather}(x) \Rightarrow \text{Man}(x) \wedge t\,(\exists\,\text{hasChild}\,.\,\text{Female}, x) =$
$\forall x\,.\; \text{HappyFather}(x) \Rightarrow \text{Man}(x) \wedge \exists y\,.\; \text{hasChild}(x, y) \wedge \text{Female}(y)$

$t\,(a{:}\text{Man} \sqcap \exists\,\text{hasChild}\,.\,\text{Female}) =$
$\text{Man}(a) \wedge (\exists y\,.\; \text{hasChild}(a, y) \wedge \text{Female}(y))$

Sintassi alternativa Lisp-like

$\top$	$\rightarrow$	`*top*`	$(\leq n\, R.C)$	$\rightarrow$	`(at-most n R C)`
$\bot$	$\rightarrow$	`*bottom*`	$\{a\}$	$\rightarrow$	`(one-of a)`
$\neg C$	$\rightarrow$	`(not C)`	R^{-}	$\rightarrow$	`(inv R)`
$C \sqcap D$	$\rightarrow$	`(and C D)`	$C \sqsubseteq D$	$\rightarrow$	`(implies C D)`
$C \sqcup D$	$\rightarrow$	`(or C D)`	$A = C$	$\rightarrow$	`(define-concept A C)`
$\exists R.C$	$\rightarrow$	`(some R C)`	$R \sqsubseteq P$	$\rightarrow$	`(implies-role R P)`
$\forall R.C$	$\rightarrow$	`(all R C)`	$\mathit{fun}(f)$	$\rightarrow$	`(functional f)`
$(\geq n\, R)$	$\rightarrow$	`(at-least n R)`	$\mathit{trans}(R)$	$\rightarrow$	`(transitive R)`
$(\leq n\, R)$	$\rightarrow$	`(at-most n R)`	$a{:}C$	$\rightarrow$	`(instance a C)`
$(\geq n\, R.C)$	$\rightarrow$	`(at-least n R C)`	$(a, b){:}R$	$\rightarrow$	`(related a b R)`