
ANNO ACCADEMICO 2024/2025

Modellazione Concettuale per il Web Semantico

Teoria

Altair's Notes



DIPARTIMENTO DI INFORMATICA

CAPITOLO 1	WEB OF DATA: I LINGUAGGI	PAGINA 5
1.1	Introduzione	5
	Perché Studiare il Semantic Web e Linked Data? — 5 • Obiettivi del Corso — 5	
1.2	Dalle Reti Semantiche alle Ontologie	6
	I Limiti della Logica Classica — 6 • Reti Semantiche — 7 • Reti Semantiche e Logica — 9 • Sistemi a Regole — 14	
CAPITOLO 2	TEST2	PAGINA 17

Premessa

Licenza

Questi appunti sono rilasciati sotto licenza Creative Commons Attribuzione 4.0 Internazionale (per maggiori informazioni consultare il link: <https://creativecommons.org/version4/>). Sono basati sulle lezioni del corso "Model-
lazione Concettuale per il Web Semantico" (2020-2021¹) dell'università di Torino: prof. Damiano. Il tutto è integrato con le slides della versione 2024-2025 del corso.



Formato utilizzato

Box di "Concetto sbagliato":

Concetto sbagliato 0.1: Testo del concetto sbagliato

Testo contenente il concetto giusto.

Box di "Corollario":

Corollario 0.0.1 Nome del corollario

Testo del corollario. Per corollario si intende una definizione minore, legata a un'altra definizione.

Box di "Definizione":

Definizione 0.0.1: Nome delle definizioni

Testo della definizione.

Box di "Domanda":

Domanda 0.1

Testo della domanda. Le domande sono spesso utilizzate per far riflettere sulle definizioni o sui concetti.

¹Perché la prof pensa che non registrare sia una cosa tanto furba e io non ho ancora sviluppato il dono dell'ubiquità per seguire due lezioni contemporaneamente.

Box di "Esempio":

Esempio 0.0.1 (Nome dell'esempio)

Testo dell'esempio. Gli esempi sono tratti dalle slides del corso.

Box di "Note":

Note:-

Testo della nota. Le note sono spesso utilizzate per chiarire concetti o per dare informazioni aggiuntive.

Box di "Osservazioni":

Osservazioni 0.0.1

Testo delle osservazioni. Le osservazioni sono spesso utilizzate per chiarire concetti o per dare informazioni aggiuntive. A differenza delle note le osservazioni sono più specifiche.

1

Web of Data: I Linguaggi

1.1 Introduzione

1.1.1 Perché Studiare il Semantic Web e Linked Data?

- Gli approcci quantitativi non sono sufficienti per domini complessi:
 - Non è possibile apprendere il comportamento giusto per ogni contesto.
 - Reattività e ragionamento.
- L'ambito della conoscenza è intrinsecamente basato su modelli:
 - Arte, media, tecnologie, etc.
- Interoperabilità dei dati:
 - Conoscenza esperta per stabilire i *mapping*.
 - Utilizzo di standard.
- Ruolo nel ragionamento in molte applicazioni:
 - Elaborazione del linguaggio naturale.
 - Question answering.
 - Chatbots.

1.1.2 Obiettivi del Corso

- Imparare a rappresentare un dominio di conoscenza con i linguaggi del Web Semantico (RDF e OWL), che permettono di implementare ontologie computazionali.
- Utilizzare strumenti di ragionamento automatico per realizzare inferenze sulla conoscenza formalizzata nelle ontologie computazionali.
- Interrogare basi di conoscenza in cui i dati sono rappresentati in un formato semantico utilizzando il linguaggio SPARQL.
- Rendere interoperabili rappresentazioni diverse (ontologie, basi di dati, fogli di calcolo) utilizzando strumenti di mapping.

1.2 Dalle Reti Semantiche alle Ontologie

1.2.1 I Limiti della Logica Classica

A partire dagli anni 60 si è sviluppata una branca dell'intelligenza artificiale specificamente orientata alla rappresentazione della conoscenza. Questo rappresenta un tentativo di superare la logica classica allora utilizzata. Il motivo è che la logica classica ha alcuni limiti importanti:

- È caratterizzata da *inadeguatezza espressiva*.
- È *monotona*.
- Presenta svantaggi dal punto di vista *computazionale*.

Definizione 1.2.1: Inadeguatezza Espressiva

Alcuni aspetti dell'inadeguatezza della logica classica possono essere attribuite a differenze con i sistemi cognitivi.

Osservazioni 1.2.1

- La logica classica è un *formalismo piatto*:
 - Tutte le affermazioni si collocano sullo stesso piano.
 - Esprime conoscenza di carattere generale e immutabile.
- Le procedure di dimostrazione sono diverse dal ragionamento umano^a.
- I *valori di verità* non sono adatti a rappresentare gli aspetti quantitativi che caratterizzano il mondo reale

^aDa premesse false discendono conseguenze vere.

Definizione 1.2.2: Monotonicità

La logica del primordine è monotona: le conoscenze inserite nel sistema logico non possono essere cancellate.

Note:-

La conoscenza può solo aumentare.

Osservazioni 1.2.2

- Le nuove conoscenze non possono contraddire quelle già presenti nel sistema.
- Impossibilità di rappresentare il cambiamento, quindi gli aspetti temporali della conoscenza.

Definizione 1.2.3: Limitazioni Computazionali

Il calcolo dei predicati è *semi-decidibile*: è possibile dimostrare con certezza cosa discende dalla base di conoscenza, ma non si ha la certezza di dimostrare ciò che non discende dalla base di conoscenza.

Corollario 1.2.1 Inefficienza nelle Procedure di Dimostrazione

Per le sue caratteristiche intrinseche (è un formalismo orientato alla rappresentazione) le procedure di dimostrazione non sono abbastanza efficienti per le applicazioni reali.

Domanda 1.1

Ma cosa possiamo dire sulle logiche non classiche?

Le logiche non classiche rilasciano alcune caratteristiche della logica classica:

- Valori di verità \Rightarrow *logiche fuzzy*, hanno valori di verità in un intervallo, nascono per superare la rappresentazione binaria dell'informazione.



Figure 1.1: Rappresentazione accurata di una logica fuzzy.

- Conoscenza certa \Rightarrow *logiche bayesiane*, incorporano il ragionamento probabilistico.
- Monotonicità \Rightarrow *default logics*, belief revision.
- Indecidibilità \Rightarrow insiemi decidibili della logica del primordine.
- Inefficienza delle dimostrazioni \Rightarrow uso di *euristiche* nella dimostrazione automatica.

Un quadro storico:

- Anni 60s-70s: reti semantiche e frames.
- Anni 70s-80s: logiche non classiche.
- Anni 80s-90s: logiche descrittive.
- Anni 2000s: ontologie computazionali.
- Dal 2010: Linked Data.

1.2.2 Reti Semantiche

Definizione 1.2.4: Reti Semantiche

Le reti semantiche sono:

- Basate su una struttura a *grafo*.
- I nodi del grafo rappresentano *concetti*.
- Gli archi tra i concetti rappresentano *relazioni tra concetti*.

Note:-

Nel "Mondo dei Blocchi" c'è un tavolo su cui è collocato un blocco verde chiamato "a" e sopra un blocco rosso chiamato "b". Una delle possibili rappresentazione è quella dell'immagine in cui sono presenti concetti, qualità

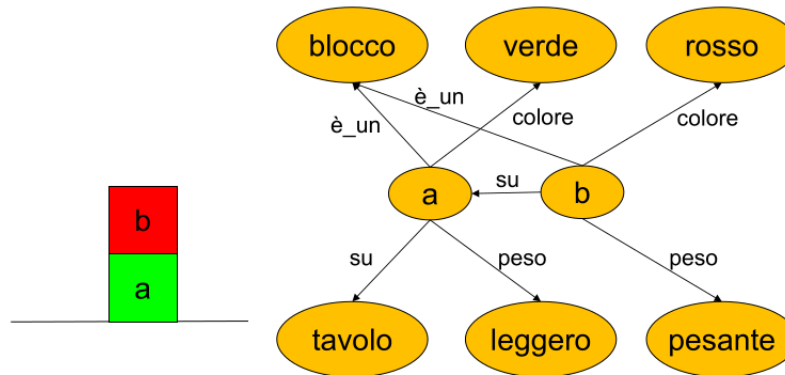


Figure 1.2: Grafo relazionale del "Mondo dei Blocchi", un dominio giocattolo utilizzato nel mondo dell'AI.

ed entità concrete. Gli archi sono varie relazioni.

Vantaggi di questa rappresentazione:

- Le informazioni relative a un nodo sono *immediatamente disponibili* dato che ogni blocco è direttamente collegato alle sue proprietà.
- Permette di rappresentare una nozione di *rilevanza* per cui dato un focus, alcune informazioni si trovano in prossimità.

Corrispondente rappresentazione logica:

- Blocco(a).
- Blocco(b).
- su(b,a).
- su(a,tavolo).
- rosso(b).
- verde(a).
- pesante(b).
- leggero(a).

Corollario 1.2.2 Ragionamento nelle Reti Semantiche

Nelle reti semantiche il ragionamento consiste nel seguire un percorso tra nodi.

Domanda 1.2

Su quale blocco si trova il blocco b?

Risposta: È sufficiente seguire l'arco etichettato come «su» nella rete, da b verso il nodo a cui punta. Seguendo ulteriormente gli archi si possono inferire relazioni indirette: il blocco b è "al di sopra" del tavolo.

1.2.3 Reti Semantiche e Logica

L'espressività delle reti semantiche corrisponde a un sottoinsieme della logica del primordine:

- Nodi = *termini*.
- Archi = *predicati*

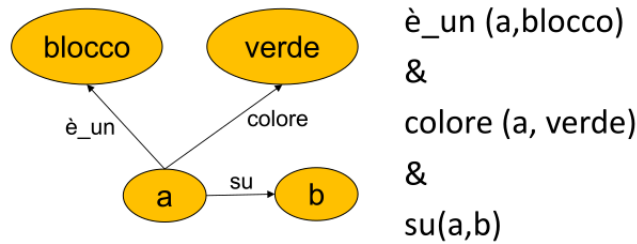


Figure 1.3: Grafo relazionale in relazione alla logica del primordine.

Note:-

Alcune relazioni possono coinvolgere più di due entità.

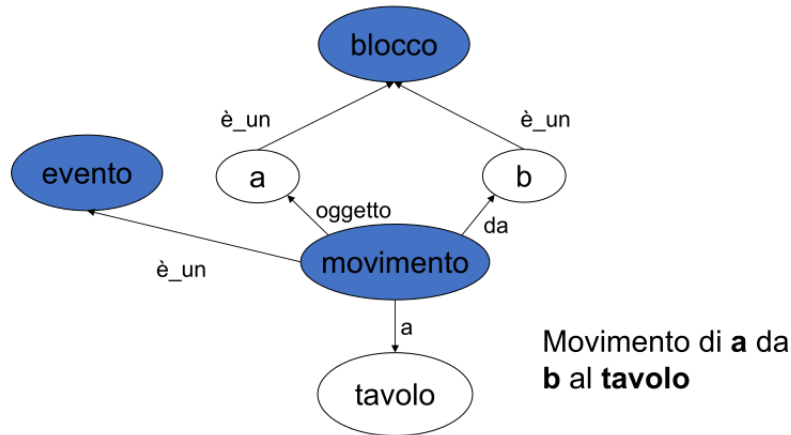


Figure 1.4: La relazione “movimento” che unisce a, b e tavolo diventa un termine cioè un nodo.

Definizione 1.2.5: Reti Semantiche Proporzionali

Le *reti semantiche proporzionali* includono nodi che rappresentano proposizioni. Usando nodi per rappresentare proposizioni è possibile introdurre una *dimensione epistemica*:

- Rappresentare credenze soggettive.
- Lo stesso sistema può rappresentare le credenze di più soggetti senza che insorgano contraddizioni.

Osservazioni 1.2.3

- Reti semantiche proporzionali posso avere l'espressività della logica del primordine una volta introdotti connettivi, variabili, quantificatori, ecc.
- Anche l'inferenza nelle reti proporzionali ha le stesse caratteristiche che nella logica del primordine.

- Soluzione: limitare l'espressività privilegiando tipi di ragionamento più comuni e computazionalmente trattabili.

Definizione 1.2.6: SNePs

Rete semantica proposizionale che incorpora alcuni elementi della teoria dei frame (Shapiro '79).

Note:-

Si tratta di una rete semantica con un *motore di ragionamento*. Permette vari tipi di ragionamento:

- Su formule.
- Su slot.
- Su percorsi.

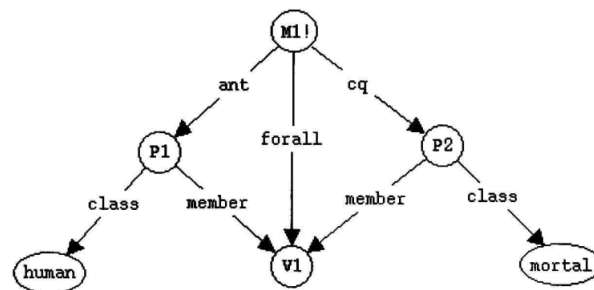


Figure 1.5: Tutti gli uomini sono mortali.

Domanda 1.3

Di cosa parla la rete?

Definizione 1.2.7: Eterogeneità

- Archi rappresentano relazioni di tipo diverso tra concetti (relazioni epistemiche vs fatti).
- Nodi rappresentano concetti di tipo diverso (blocco A, blocco)

Note:-

Alcuni tipi di nodi e di archi sono particolarmente importanti.

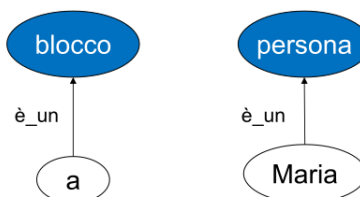


Figure 1.6: I nodi colorati rappresentano tipi di entità (classi), i nodi bianchi rappresentano entità singole (individui).

Domanda 1.4

"What ISA is and isn't?"

- Gli archi "è un" (IS-A o ISA) hanno un significato diverso se collegano due classi oppure un individuo a una classe.
- Brachman ('83) propone di distinguere i due tipi di relazioni: "What isa is and isn't":
 - *Archi ISA*: appartenenza di una classe a una sottoclasse (transitiva).
 - *Archi instance_of*: appartenenza di un individuo a una classe.

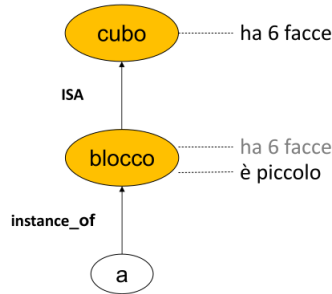


Figure 1.7: ISA e instance_of.

Definizione 1.2.8: Tassonomia

Utilizzando la relazione ISA è possibile esprimere conoscenza di tipo tassonomico. Si utilizza un ragionamento classificatorio:

- Y è una sottoclasse di Z? \Rightarrow archi ISA.
- x appartiene alla classe Y? \Rightarrow archi instance_of.

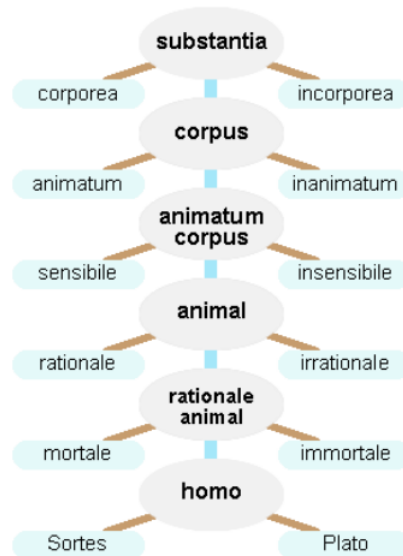


Figure 1.8: Rappresentazione che deriva dalla filosofia aristotelica, confluita nella scolastica, in cui si raffigurano i principali concetti del pensiero dalla sostanza all'uomo.

Note:-

Ogni livello *eredità* le caratteristiche del livello superiore e le caratteristiche non si sovrappongono.

Vantaggi dell'ereditarietà:

- E' possibile avere una rappresentazione meno ridondante facendo una semplice assunzione:
 - Una classe eredita le proprietà delle classi di cui è sottoclasse.
 - Se i cubi hanno sei facce e i dadi sono una sottoclasse dei cubi, allora anche i dadi hanno sei facce.
- Le sottoclassi hanno proprietà più specifiche delle classi.
- Seguendo il percorso degli archi ISA e instance_of e diventa possibile ragionare sulle proprietà di un individuo/classe.

Note:-

Ma esistono eccezioni, per esempio i pinguini non volano, ma non per questo smettono di essere uccelli. Allo stesso tempo può essere falso per un individuo all'interno di una sottoclasse.

Definizione 1.2.9: Default Rules

Per gestire le eccezioni è necessario rilasciare la proprietà della monotonicità (la conoscenza non diminuisce mai). Un esempio è la Default Logic: assunzioni che possono essere cancellate quando sopravviene nuova conoscenza.

Osservazioni 1.2.4

- Il trattamento delle eccezioni nelle reti semantiche si basa sul principio che le conoscenze specificate localmente a un certo nodo prevalgono su quelle ereditate.
- Un corollario è che le conoscenze che comportano meno passi di inferenza prevalgono su quelle che ne comportano di più.
- Questo principio però non permette di scongiurare tutti gli inconvenienti.

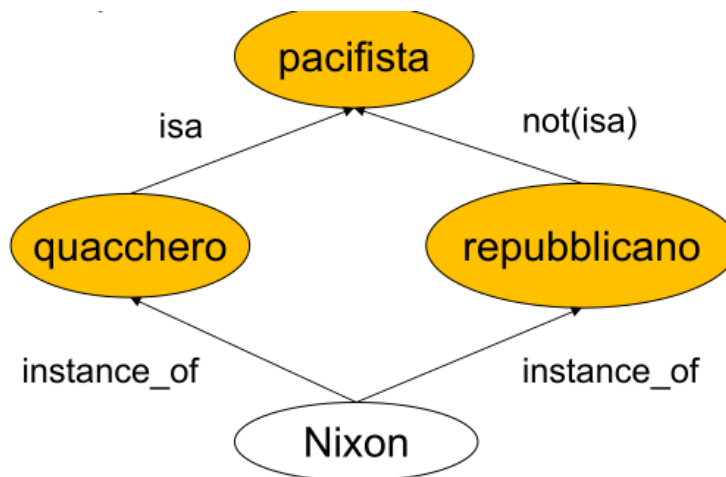


Figure 1.9: Nixon è pacifista o guerrafondaio? In quanto quacchero, lo è; in quanto repubblicano, non lo è (Reiter and Criscuolo, On Interacting Defaults, 1981).

Note:-

Nemmeno la logica dei default risolve il problema, ciascun default blocca l'altro.

Definizione 1.2.10: Frame Theory

Evoluzione delle reti semantiche finalizzata a rappresentare la conoscenza di tipo stereotipato. Conoscenza di sfondo utile per alcune applicazioni come la *visione artificiale* o *l'elaborazione del linguaggio*.

Definizione 1.2.11: Frame

Un *frame* è una struttura dati per la rappresentazione situazioni stereotipate, come trovarsi in una certa tipo di soggiorno o andare al compleanno di un bambino partito. I livelli superiori del telaio sono fissi, e rappresentano cose che sono sempre vere riguardo le situazioni presunte. I livelli inferiori hanno molti terminali: slot che devono essere riempiti da istanze o dati specifici.

Corollario 1.2.3 Slot

Ogni terminale (slot) può specificare le condizioni che Gli incarichi devono soddisfarsi. (Gli incarichi sono solitamente telai ausiliari più piccoli). Le condizioni possono richiedere un terminale l'incarico di essere una persona, un oggetto ora puntatore a un subframe di un certo tipo. I terminali di un frame sono solitamente riempiti da assegnazioni predefinite. Collezioni di frame collegati sono uniti in frame systems.

Struttura di un frame:

- Identificativo.
- Slot: permettono di creare una tassonomia di frame.
- Slot generici: i valori degli slot sono vincolati a un certo tipo, però il contenuto di uno slot può puntare a un altro frame.
- Procedure per il calcolo automatico dei valori.
- Valori predefiniti.

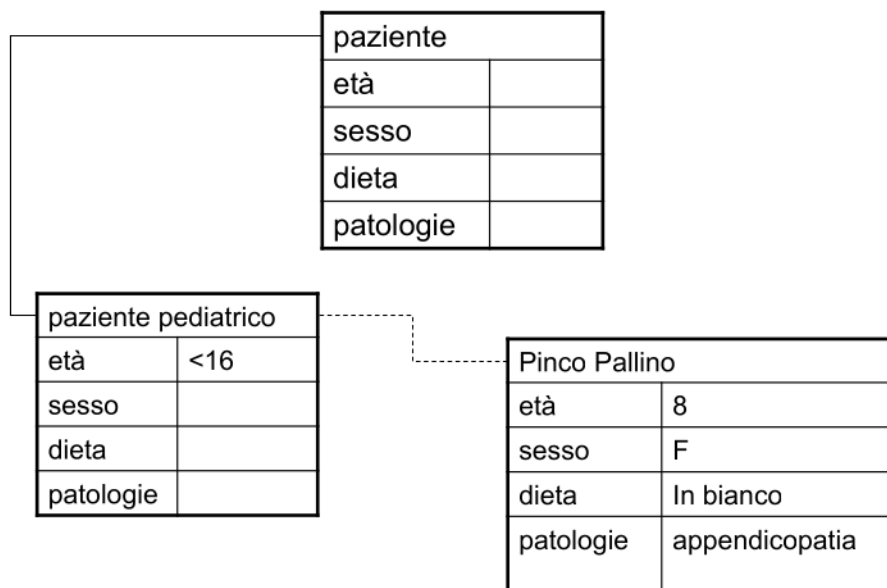


Figure 1.10: Esempio di tassonomia.

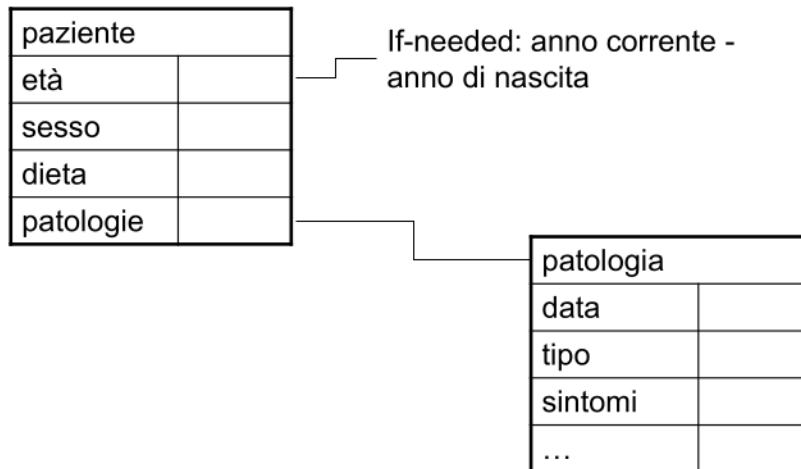


Figure 1.11: Esempio di collegamenti non tassonomici.

Definizione 1.2.12: Regole di Produzione

Conoscenza di tipo condizione-azione. Regole IF - THEN (anche dette produzioni):

- *La parte sinistra*: condizioni per l'applicazione delle regole.
- *La parte destra*: azioni che sono effettuate se la condizione è soddisfatta.

Osservazioni 1.2.5

- Conoscenza dichiarativa:
 - IF allatta(animale,prole) THEN isa(animale,mammifero).
 - IF febbre>38,5 and tosse THEN malattia_da_raffreddamento.
- Conoscenza procedurale:
 - IF intensità(sibilo)>t THEN chiudi(valvola,5mm).

1.2.4 Sistemi a Regole

Componenti:

- Base delle regole.
- Memoria di lavoro.
- Interprete delle regole: motore inferenziale.

L'interprete delle regole esegue questo ciclo:

1. Confronto (match) tra fatti (nella memoria di lavoro) e antecedente delle regole.
2. Risoluzione di conflitti (più regole applicabili).
3. Esecuzione (e aggiunta di nuovi fatti alla memoria di lavoro).

Conflitti di applicazione:

- *Refrattarietà*: la stessa regola non viene applicata più volte.
- *Recenza*: vengono privilegiate le regole che si applicano ai fatti più recenti.
- *Specificità*: vengono applicate le regole più specifiche.
- *Pesatura*: assegnamento di importanza (peso) a priori alle regole o ai fatti.

Note:-

Un passo delicato è il confronto tra la memoria di lavoro e l'antecedente di tutte le regole,

Definizione 1.2.13: Mycin

Sistema esperto sviluppato a Stanford nei primi anni 70s (Shortliffe, '75). I fatti sono associati alle probabilità.

PREMISE: (AND

(SAME CNTXT GRAM GRAMNEG)

(SAME CNTXT MORPH ROD)

(SAME CNTXT AIR ANAEROBIC))

ACTION: (CONCLUDE CNTXT IDENTITY BACTEROIDES TALLY .6)

- IF: 1) The site of the culture is blood, and 2) The gram stain of the organism is gramneg, and 3) The morphology of the organism is rod, and 4) The portal of entry of the organism is udne, and 5) The patient has not had a genito-urinary manipulative procedure, and 6) Cystitis is not a problem for which the patient has been treated THEN: There is *suggestive evidence* (.6) that the identity of the organism is e.coli

Figure 1.12: Esempio di Mycin.

2

Test2

