

Progetto Finale:

Modelli semantici dei dati in domini applicativi specifici

Roberto Basili, Marco Pennacchiotti

Corso di Fondamenti di Informatica

a.a. 2003-2004

Definizioni Generali e Regole d'esame

Il progetto fa parte della valutazione globale del Corso di **Fondamenti di Informatica** sia per gli studenti (*nuovo ordinamento*) che ricevono dal Corso 5 crediti formativi, sia per quelli (vecchio ordinamento) che invece accederanno a 10 crediti. Questa distinzione verrà utilizzata in ognuna delle descrizioni delle proposte di progetto per distinguerne la complessità relativa alle diverse fasi realizzative richieste agli studenti: fasi di crescente complessità daranno diritto ad un numero di crediti maggiori e saranno quindi consigliate alle diverse tipologie di studenti del corso. Il progetto deve essere completato e consegnato prima della verbalizzazione dell'esame di cui costituisce (insieme al Test Finale¹) parte integrante.

L'obiettivo dei progetti è quello di costruire una sistematica descrizione di un ambito di conoscenza (per esempio un insieme di settori filosofici, un insieme di generi musicali) ed utilizzare la tecnologia delle basi di dati o della rappresentazione della conoscenza per la realizzazione di un modello di dati che ne esprima le caratteristiche semantiche di rilievo, cioè:

- Le classi principali in cui i dati e le informazioni di tale dominio si organizzano
- Le relazioni di *sottoclasse-di* che esistono tra le classi definite, e quindi la gerarchia eventuale che da esse risulta
- Gli attributi e le proprietà che caratterizzano tutti gli oggetti di una classe secondo la gerarchia definita
- La appartenenza di un numero non banale di entità (o *istanze*) alle classi della gerarchia

La sistematica descrizione che risulta dalla definizione data di sopra sarà d'ora in poi definita come *modello semantico del dominio*. Tipico esempio di un tale modello è la classificazione linguistica che si può imporre ad un insieme di verbi, nomi ed aggettivi per descriverne le proprietà morfosintattiche quali: genere, numero, modo, transitività, ... Il risultato atteso da un progetto è quindi quello di una realizzazione di tale modello, non solo dal punto di vista della sua definizione, ma anche una sua implementazione al computer mediante una opportuna piattaforma per la gestione dei dati. Nel caso in cui il modello non usasse alcuna modalità di inferenza automatica lo strumento d'elezione dovrebbe essere un sistema di gestione di una base di dati, in particolare *Microsoft Access*, trattato a lezione. Per i modelli che richiedano l'uso della nozione di sottoclasse ed ereditarietà delle proprietà allora lo studente potrà fare riferimento alla piattaforma per l'editing e la manutenzione di ontologie quali il tool *Protege 2000*, trattato a lezione. In

¹Nel sito del Corso, URL www.uniroma2.it/didattica/FondInform/index.html sono riportati esempi di Domande del Test Finale

ogni caso la validita' del modello dovra' poter essere verificato, mostrando che la implementazione fornisce risposte corrette ad un insieme prefissato di interrogazioni (*query*). La risorsa prodotta infatti dovra' rispondere correttamente a tutte le *query* esplicitamente riportate nel testo del progetto.

Ogni progetto e' caratterizzato dalle seguenti fasi:

- Fase di *studio del problema*, in cui il dominio in esame verra' esaminato in dettaglio prima della realizzazione pratica della modello semantico. Il risultato di questa fase e' la definizione del problema (cioe' la definizione dei limiti del dominio che si intende descrivere), la definizione del materiale (o dei dati) raccolti come in *input* e la definizione del risultato atteso in *output*. Ad es. l'insieme delle interrogazioni a cui si intende dare risposta e' un *output* atteso.
- Fase di *realizzazione* in cui verranno utilizzate delle piattaforme e degli ambienti *software* per la costruzione della soluzione al problema. Tipici esempi di tali strumenti software sono quindi *Microsoft Access* (per lo sviluppo delle basi di dati) e *Protege 2000* (per lo sviluppo di ontologie).
- Fase di *analisi dei risultati*, in cui, dopo aver realizzato il modello semantico ed avere inserito alcune istanze relative al dominio considerato, lo studente descrivera' la qualita' (accuratezza della descrizione e livello di sintesi raggiunto nel modello) della risorsa costruita. Questa analisi potra' essere effettuata rispetto ad un insieme di documenti preso a riferimento. Ad esempio, nella costruzione di una ontologia per le discipline filosofiche, lo studente potra' descrivere la qualita' della ontologia ottenuta rispetto al materiale critico (come le descrizioni al sito ... utilizzate durante la lezione/esercitazione). Tipici criteri saranno qui la discussione del numero di concetti della gerarchia ottenuta (e delle corrispondenti proprieta') rispetto a quelli utilizzati dalla critica di riferimento. Un'altra modalita' di valutazione e' quella delle interrogazioni (*query*) tipiche che verranno suggerite nella definizione dei singoli progetti.
- Fase di *documentazione*, in cui lo studente potra' descrivere le fasi precedenti in un documento (5-10 pagine al massimo) da allegare alla consegna della risorsa.

Lo studente dovra' quindi consegnare una relazione contenete sezioni distinte per le seguenti attivita'

- Fase di *studio del problema*
- Progettazione concettuale
- Fase di *realizzazione*
- Fase di *analisi dei risultati*
- Test di risposta alle domande

Si vuole osservare che ogni progetto, partendo dalle specifiche (parziali) fornite da questo documento, puo' essere sviluppato secondo scelte progettuali diverse a discrezione dello studente. La relazione dovra' quindi descrivere le assunzioni aggiuntive considerate da ogni specifico progetto.

Un progetto puo' essere svolto da un gruppo formato al massimo da due studenti. Ogni gruppo dovra' consegnare una sola relazione. La data finale di consegna dei progetti e' entro tre giorni dalla data della discussione orale del progetto stesso

1 Progetto 1: Definizione di un modello morfosintattico di un lessico per la lingua italiana.

Si progetti ed implementi in un ambiente per la progettazione semantica dei dati (cioè MS Access o Protege) un lessico della lingua italiana per la rappresentazione dei fenomeni morfosintattici. La gerarchia delle classi lessicali da rappresentare dovrà descrivere almeno le seguenti proprietà:

- caratteristiche grammaticali dei lemmi (ad es. separare sostantivi da verbi ed aggettivi)
- caratteristiche morfosintattiche quali genere e numero
- caratteristiche grammaticali quali classi verbali (ad es. transitivi vs. intransitivi) e classi aspettuali

Il modello risultante dovrebbe poter facilitare la introduzione di nuove entrate lessicali in un dizionario elettronico. In particolare, permettere ad esempio di inserire le sole radici dei verbi regolari e di "ereditare" dalle proprietà della coniugazione (una classe verbale possibile quindi) le desinenze dei vari tempi verbali.

E' necessario ovviamente popolare la base di dati o l'ontologia risultante con un insieme di istanze rappresentativo dei diversi casi e rendere quindi la risposta alle diverse interrogazioni (*query*), proposte di seguito, non banale (cioè ridotta ad un unico elemento).

Il sistema dovrà rispondere a domande quali:

- *Quali sono i verbi transitivi rappresentati?*
- *Quali sono i nomi esclusivamente plurali?*
- *Quali sono le proprietà dei verbi transitivi rappresentati?*
- *Quali sono le desinenze possibili di una certa classe verbale?*
- *Quali sono i nomi la cui forma maschile e femminile è identica?*

Un esempio di descrizione di un modello possibile (incentrato sulla categoria dei verbi) può essere trovato al sito: URL:
<http://www.uni-stuttgart.de/lingrom/stein/pubs/pdf/stein99a.pdf>.

1.1 Richieste del Progetto

La realizzazione del progetto dovrà prevedere almeno la analisi completa di **una** tra le seguenti categorie lessicali

- nomi
- verbi
- aggettivi

Per la categoria scelta si dovranno specificare chiaramente gli scopi della ontologia (per es. descrivere le classi desinenziali per poter automaticamente calcolare le forme flesse a partire dalle radici dei lemmi) ed un insieme di domande (*query*) che sin dall'inizio ci si propone di soddisfare in ambiente Protege o, in alternativa, in MS Access.

1.2 Crediti

I crediti forniti da questo progetto si dividono nelle seguenti classi:

- 5 crediti per lo sviluppo di una sola categoria lessicale a scelta.
- 10 crediti per lo sviluppo di due o piu' classi.
-

2 Progetto 2: Definizione di un modello semantico di un lessico.

Si progetti ed implementi in un ambiente per la progettazione semantica dei dati (cioe' MS Access o Protege) un lessico semantico delle parole della lingua italiana. Intendiamo come lessico semantico una rete di concetti con ereditarieta' le cui classi sono dei concetti rappresentati dalla lingua in elementi lessicali nominali. La gerarchia delle classi lessicali da rappresentare dovra' descrivere almeno le seguenti proprieta':

- la nozione di *senso* per ogni lemma. Similmente ai dizionari un *senso* sara' dato da una specifica accezione del lemma.
- caratteristiche definitorie dei sensi (ad es. una breve definizione simile a quelle adottate dai dizionari)
- relazioni tra i concetti dati dai singoli sensi (per esempio relazione *is_a_kind_of* tra sensi di diversa generalita', come "*marcia*" *is_a_kind_of* (*e' un tipo di*) "*cammino*").

Il modello risultante dovrebbe poter facilitare la introduzione di nuovi sensi lessicali come in un dizionario elettronico. In particolare, permettere ad esempio di inserire un nuovo senso (per es. "*corsa*") in una specifica posizione della gerarchia di concetti (per esempio come tipo di "*cammino*") ed "*ereditare*" automaticamente le proprieta' note per i sensi piu' generali (cioe' "*cammino*").

E' necessario ovviamente popolare la base di dati o l'ontologia risultante con un insieme di istanze rappresentativo dei diversi casi e rendere quindi la risposta alle diverse interrogazioni (*query*), proposte di seguito, non banale (cioe' ridotta ad un unico elemento).

Il sistema dovra' rispondere a domande quali:

- *Quali sensi per le "AZIONI" sono rappresentati?*
- *Quali nomi rappresentano sensi per le "AZIONI"?*
- *Quali nomi rappresentano "LUOGHI" e "ASTRAZIONI"?* (ad es. la prola *scuola*)

Un esempio di rete semantica per l'inglese e' Wordnet che puo' essere usato come fonte di ispirazione. URL:

<http://www.cogsci.princeton.edu/wn/>.

2.1 Richieste del Progetto

La realizzazione del progetto dovra' prevedere almeno la analisi completa di **un** testo breve (novella o poesia) e la definizione dei sensi per tutti i nomi utilizzati nel testo. Per tale insieme di lemmi si dovranno specificare chiaramente gli scopi della ontologia dei sensi (per es. descrivere le caratteristiche semantiche utili allo scopo comunicativo dell'autore) ed un insieme di domande (*query*) che sin dall'inizio ci si propone di soddisfare in ambiente Protege o, in alternativa, in MS Access.

2.2 Crediti

I crediti forniti da questo progetto si dividono nelle seguenti classi:

- 5 crediti per lo sviluppo della sola categoria dei nomi.
- 10 crediti per lo sviluppo della ontologia dei nomi e dei verbi.
-

3 Progetto 3: Beni culturali

Si progetti ed implementi in un ambiente per la progettazione semantica dei dati (cioe' MS Access o Protege) un modello dei generi letterari contestualizzati nei diversi periodi storici e contenente informazioni sugli autori con riferimenti alle proprie opere e al proprio inquadramento storico e di genere (Suggerimento: creare una gerarchia dei generi, una delle opere, una dei periodi storici ed una degli autori e creare relazioni tra di essi). Si popoli il modello con un numero di istanze sufficienti a mostrarne l'adeguatezza, e rendere quindi la risposta alle diverse interrogazioni (*query*), proposte di seguito, non banale (cioe' ridotta ad un unico elemento).

Il sistema dovra' rispondere a domande quali:

- *Quali sono i generi letterari rappresentati?*
- *Quali sono gli autori di un dato genere letterari?*
- *Quali sono le opere pubblicate in un dato periodo storico?*
- *Quali sono le opere di un dato genere letterario?*
- *Quali sono i nomi la cui forma maschile e femminile e' identica?*

Un esempio di descrizione di un modello possibile puo' essere trovato al sito: URL: http://it.wikipedia.org/wiki/Generi_di_letteratura.

Si consiglia in ogni caso di effettuare ulteriori ricerche sul web.

3.1 Richieste del Progetto

La realizzazione del progetto dovra' prevedere almeno la analisi completa di **due** aspetti del modello tra:

- generi
- periodi storici
- autori
- opere

3.2 Crediti

I crediti forniti da questo progetto si dividono nelle seguenti classi:

- 5 crediti per lo sviluppo di due soli aspetti del modello.
- 10 crediti per lo sviluppo del modello completo.

4 Progetto 4: Classificazione Musicale dei Generi

Si progetti ed implementi in un ambiente per la progettazione semantica dei dati (cioe' MS Access o Protege) un modello dei generi musicali legati alla musica popolare (ad es. blues, classica, disco, jazz, pop, rock). L'interesse di questo progetto e' nella definizione dei generi in base a particolari caratteristiche musicali e tecniche. Si supponga che l'utilizzo finale del modello sia di supporto ad un sistema per la classificazione automatica dei brani musicali nei diversi generi musicali: il modello dovrebbe essere in grado di fornire criteri efficaci al fine di individuare il genere relativo ad un certo brano fornito in ingresso. Prima il brano (ad es. un file MP3 o MIDI) dovrebbe poter essere elaborato con tecniche di computer music per la estrazione di caratteristiche musicali (per es. tipo di strumenti impiegati o ritmo del brano). Infine tali caratteristiche in accordo con la ontologia dovrebbero essere usate per deciderne il genere. Questa tipica applicazione Web potrebbe estendere un software di tipo *player* e fornire criteri sistematici per la classificazione in genere.

Potanno ad esempio essere considerate come caratteristiche significative gli strumenti presenti in un brano, la sua metrica, l'estensione delle note o gli intervalli melodici.

Si popoli il modello con un numero di istanze sufficienti a mostrarne l'adeguatezza, e rendere quindi la risposta alle diverse interrogazioni (*query*), proposte di seguito, non banale (cioe' ridotta ad un unico elemento).

Il sistema dovra' rispondere a domande quali:

- *Quali sono i generi musicali rappresentati?*
- *Quali sono le caratteristiche principali di un dato genere musicale?*
- *Quali sono gli strumenti utilizzati in un dato genere musicale?*
- *Quali sono le caratteristiche tipiche del "jazz" e del "blues"?*

4.1 Richieste del Progetto

4.2 Crediti

I crediti forniti da questo progetto si dividono nelle seguenti classi:

- 5 per lo sviluppo del modello completo.
- 10 nel caso in cui si popoli la ontologia con un numero di brani in formato MIDI o MP3 maggiore di 100, dotandoli di relative classificazioni

5 Progetto 5: Generi Musicali

Si progetti ed implementi in un ambiente per la progettazione semantica dei dati (cioe' MS Access o Protege) un modello dei generi musicali e degli artisti, contestualizzati in diversi periodi storici. Ogni genere (o sotto-genere) deve essere descritto da sua proprieta' caratteristiche tecniche ed espressive, ed avere un riferimento al (ai) periodo (i) storico(i). Si rappresenti inoltre nel modello una gerarchia di strumenti musicali, con riferimenti ai diversi generi. Si popoli il modello con un numero di istanze sufficienti a mostrarne l'adeguatezza, e rendere quindi la risposta alle diverse interrogazioni (*query*), proposte di seguito, non banale (cioe' ridotta ad un unico elemento).

Il sistema dovra' rispondere a domande quali:

- *Quali sono i generi musicali rappresentati?*
- *Quali sono gli artisti di un dato genere musicale?*
- *Quali sono gli strumenti utilizzati in un dato genere musicale?*
- *Quali sono i periodi storici di un dato genere musicale?*

Un esempio di descrizione di un modello possibile puo' essere trovato al sito: URL: http://it.wikipedia.org/wiki/Generi_musicali.

Si consiglia in ogni caso di effettuare ulteriori ricerche sul web.

5.1 Richieste del Progetto

La realizzazione del progetto dovra' prevedere almeno la analisi completa di **due** aspetti del modello tra:

- generi
- periodi storici
- artisti
- strumenti

5.2 Crediti

I crediti forniti da questo progetto si dividono nelle seguenti classi:

- 5 crediti per lo sviluppo di due soli aspetti del modello.
- 10 crediti per lo sviluppo del modello completo.

6 Progetto 6: Discipline Filosofiche

Si progetti ed implementi in un ambiente per la progettazione semantica dei dati (cioe' MS Access o Protege) un modello della Storia della Filosofia che rappresenti una gerarchia delle diverse parti della filosofia (metafisica, epistemologia,ecc.). Si rappresentino inoltre i filosofi, collocati nel loro periodo storico, e con riferimenti alle diverse parti della filosofia cui hanno contribuito e con quali opere. Si popoli il modello con un numero di istanze sufficienti a mostrarne l'adeguatezza, e rendere quindi la risposta alle diverse interrogazioni (*query*), proposte di seguito,non banale (cioe' ridotta ad un unico elemento).

Il sistema dovra' rispondere a domande quali:

- *Quali sono le diverse parti della filosofia rappresentate?*
- *Quali sono i filosofi di un dato periodo storico?*
- *Quali sono le opere di un dato filosofo?*
- *Quali sono i periodi storici di una data parte della filosofia?*

Un esempio di descrizione di un modello possibile puo' essere trovato al sito: URL: http://www.culturanuova.net/filosofia/filo_temi.php.

Si consiglia in ogni caso di effettuare ulteriori ricerche sul web.

6.1 Richieste del Progetto

La realizzazione del progetto dovra' prevedere almeno la analisi completa di **due** aspetti del modello tra:

- parti della filosofia
- periodi storici
- filosofi
- opere

6.2 Crediti

I crediti forniti da questo progetto si dividono nelle seguenti classi:

- 5 crediti per lo sviluppo di due soli aspetti del modello.
- 10 crediti per lo sviluppo del modello completo.

7 Progetto 7: Scelta Facoltativa del dominio di applicazione

Si consideri un dominio specifico di interesse. Si definisca ed implementi quindi in un ambiente per la progettazione semantica dei dati (cioe' MS Access o Protege) il relativo modello, descrivendo il dominio scelto e giustificando esaurientemente le scelte progettuali ed implementative. In base al previsto utilizzo del modello si elaborino alcune interrogazioni (*query*), e si popoli il modello con un numero di istanze sufficienti a mostrarne l'adeguatezza. Rendere la risposta alle diverse query non banale.

7.1 Crediti

I crediti forniti da questo progetto sono 5. Verranno di volta in volta considerati, in base alla qualita' del progetto stesso, eventuali progetti da 10 crediti.