

Glossari in ambienti didattici ipertestuali

Luigi Colazzo & Marco Costantino

Universita' di Trento
Dipartimento di Informatica e Studi Aziendali
Via Inama 1, 38100, Trento
Tel. 0461 - 882144
Fax. 0461 - 882124
E.M. colazzo@cs.unitn.it

Abstract

In questo lavoro presentiamo una classe di oggetti chiamata G_lossario. Le occorrenze della classe G_lossario (Glossari) sono ipertesti specializzati per un ambiente ipertestuale didattico. Un oggetto della classe G_lossario e' un'aggregazione di oggetti. Tra le classi di oggetti che compongono un Glossario e' rilevante la classe D_efinizione (Definizioni). Un Glossario puo' contenere zero o piu' oggetti della classe D_efinizione. Una occorrenza di D_efinizione e' un oggetto specializzato che permette di visualizzare le definizioni del glossario. Un glossario risponde ai messaggi provenienti da altri oggetti ipertestuali presenti nell'ambiente, materializzando nell'ipertesto chiamante copie di oggetti della classe D_efinizione. Le copie di Definizioni ricevute da un ipertesto che le ha richieste diventano componenti dell'ipertesto richiedente. Poiche' le occorrenze della classe D_efinizioni sono frammenti di un ipertesto, esse possono contenere anche ancora che individuano punti di partenza di link, sicche' le copie di una occorrenza della classe D_efinizione conservano inalterate le proprie ancora quando sono incorporate in un ipertesto chiamante. La classe G_lossario e' un esempio di cio' che intendiamo per materiali didattici ipertestuali.

Lavoro Completo
Ipertesti e sistemi multimediali

Premessa

Tra le opere che costituiscono una letteratura alcune svolgono un ruolo strumentale all'attività di lettura come i dizionari, i glossari, gli atlanti, le tavole di concordanza, le cronologie, ecc. Questa categoria di opere ha un ruolo speciale nell'universo letterario. In una interpretazione metaforica esse sono porte di comunicazione tra corpus di conoscenze differenti e tra esperienze di lettura eterogenee. Così, infatti, l'atlante che vi ha accompagnato nella lettura di Giulio Verne è identico (da un punto di vista morfologico) a quello che vi accompagna nella lettura dell'opera di Darwin.

In questo lavoro prendiamo in considerazione una specifica classe di opere letterarie: i glossari.

I glossari sono opere letterarie di consultazione (solitamente allegate ad un'opera che tratti argomenti tipici di una subcultura¹ in cui si fornisce una definizione per ogni termine specialistico. Alle volte i glossari assumono la consistenza di opere letterarie autonome e, in questo caso, sono commercialmente noti come dizionari. I glossari sono estremamente utili nelle fasi iniziali dello studio di un nuovo corpus di conoscenze. Gli autori hanno osservato empiricamente, durante la loro attività di insegnamento di un corso universitario di informatica di base, che ad ogni incremento di conoscenza dei termini specialistici della disciplina corrisponde un incremento nella capacità di studio autonomo da parte degli studenti. Nonostante la loro indubbia utilità occorre però una certo rigore intellettuale per il loro uso produttivo. La ragione apparente è che, in genere, i dizionari e i glossari sono opere voluminose e spesso la pigrizia ha il sopravvento sulla sete di conoscenza. I lettori non avvezzi ad un minimo di rigore tendono a costruire ipotesi temporanee di significato per i termini sconosciuti sperando che nei passi successivi la semantica dei termini ignoti diventi comprensibile dal contesto del discorso. Questa scorciatoia intellettuale, se può avere in certe limitate circostanze un effetto positivo, nella maggior parte dei casi si rivela: (a) una perdita di tempo, perché il contesto non chiarisce quasi mai ciò che è dato per scontato da chi utilizza un termine specialistico; (b) una fonte di confusione, perché le ipotesi incomplete sulla semantica dei termini producono bislacche deduzioni; (c) un ostacolo alla comprensione del testo. Se si osserva però più attentamente il fenomeno, si nota che non è la sola pigrizia ad indurre molti lettori a non consultare un glossario in presenza di un termine ignoto. Infatti se così fosse la disponibilità di un glossario elettronico indurrebbe nei lettori di questi testi un maggior uso dei glossari stessi. Il che però non accade, o almeno non sembra accadere da ciò che empiricamente osserviamo. Ci sembra infatti che la ragione di uno scarso uso di questo tipo di opere sia di una natura più sottile. Infatti siamo convinti che ciò che disturba il lettore è la perdita del contesto. Nel caso di un testo a stampa, il processo di consultazione può essere schematizzato come segue: (a) un termine è riconosciuto come sconosciuto nella lettura del testo X; (b) si sospende la lettura di X; (c) si apre il dizionario; (d) si cerca il termine ignoto; (e) si legge la definizione; (f) si riprende la lettura del testo X dal punto in cui ci si era interrotti. Ciò che accade è che il lettore passa dal contesto del testo X allo speciale contesto del dizionario. Cioè passa da un contesto nella maggior parte dei casi, organizzato come una narrazione (il contesto del testo X) ad un contesto organizzato come un elenco (il contesto del dizionario) e viceversa. I glossari computerizzati molto spesso simulano il medesimo meccanismo di interazione producendo un'analoga perdita di contesto. Ad esempio, si consideri la seguente interazione: (a) richiamare il programma Glossario; (b) attivare l'opzione ricerca termine; (c) comporre la stringa cercata nell'apposito box di dialogo; (d) dare la conferma; (e) la finestra originale è sostituita da una finestra sovrapposta contenente il testo della definizione; (f) leggere la definizione; (g) chiudere il glossario; (h) tornare al testo di partenza. Nel prototipo che presentiamo questi aspetti del problema ci hanno suggerito di usare una nuova modalità di interazione tra glossario e lettore di un ipertesto. Nel nostro approccio un oggetto della classe *G_lossario* è un dispensatore di copie di oggetti della classe *D_efinizioni*. Il glossario cioè si comporta come un *server* di definizioni. Le applicazioni clienti diventano proprietarie delle copie

¹Il termine subcultura è qui usato nell'accezione specialistica della sociologia e significa cultura di un gruppo sociale separato e non cultura di basso livello come spesso avviene nell'uso comune

ed ereditano così tutti i link stabiliti dall'autore della definizione. Discutiamo in maggior dettaglio questi aspetti nella sezione 2.

I glossari sono opere letterarie organizzate secondo una specifica morfologia. E' conveniente rispettare questa morfologia nel processo di traduzione da opera cartacea ad opera elettronica. Infatti questa scelta aumenta le possibilità di un rapido apprendimento delle funzioni e delle modalità d'uso di un glossario elettronico da parte degli utenti. Da questo punto di vista si può dire che i glossari elettronici devono essere " *banali*", ovvero:

- le definizioni devono poter essere lette in sequenza diretta o inversa;
- si deve poter accedere direttamente a una definizione tramite un indice alfabetico;
- si deve poter accedere direttamente a una definizione immettendo una stringa di ricerca.

Questa necessità di " *banalità*" e' rispettata nel prototipo, poiché: un oggetto della classe *G_lossario* può essere sfogliato, permette di accedere ad una definizione sia mediante una consultazione di un indice, sia impostando una stringa di ricerca. Questi aspetti del programma saranno messi in evidenza nella sezione 3. Il requisito di un isomorfismo metaforico tra strumento elettronico e strumento cartaceo non implica però, la necessità di rinunciare agli aspetti innovativi forniti dalla disponibilità di un processore artificiale, appiattendolo i glossari elettronici alle loro prestazioni banali. Infatti riteniamo che uno strumento elettronico debba fornire funzionalità non ottenibili con la tecnologia della carta. La funzionalità di ottenere le definizioni senza uscire dal contesto di ciò che si sta leggendo ci sembra essere una di queste funzionalità aggiuntive.

L'obiettivo applicativo del lavoro e' quello di dimostrare la fattibilità di progetti di ipertesti che siano al contempo pubblici e privati. Ciò e' essenziale in un ambiente applicativo dedicato a favorire processi di apprendimento attraverso l'utilizzo di tecnologie ipertestuali. Un ipertesto infatti può essere usato in modi diversi in un ambiente didattico. Può essere utilizzato per studiare, può essere utilizzato per cercare, può essere usato per esporre, e infine può essere utilizzato per costruire altre opere ipertestuali. [Cola95]. In tutti questi casi un'opera ipertestuale deve esibire comportamenti adeguati al compito applicativo per cui e' stata progettata ma deve anche garantire la possibilità di comunicazione tra i diversi soggetti coinvolti nel processo e una intercambiabilità di ruoli la quale costituisce la più importante potenzialità didattica delle tecnologie ipertestuali. J.J Leggett e J.L. Schanase [Legg94] suggeriscono che il problema della condivisione di ipertesti possa essere affrontato secondo due punti di vista alternativi. Il primo comporta lo scambio tra i soggetti cooperanti di opere organizzate come ipertesti e ciò implica il problema della traduzione degli ipertesti scambiati da un sistema ipertestuale ad un'altro qualora i due soggetti non usino il medesimo software di base. Il secondo, definito *hypermedia in the large*, prevede invece l'accesso condiviso ad una rete di dati distribuita su più macchine dislocate in più siti. Si può però immaginare che la cooperazione possa essere realizzata secondo una modalità alternativa, ovvero secondo un modello che potremmo definire di *distribuzione & diffusione*. In questo approccio in *the large*, un opera ipermediale si comporta anche come un distributore di componenti riutilizzabili in opere ipertestuali realizzate con differenti sistemi di sviluppo. Se i componenti distribuiti sono ben standardizzati, gli ipertesti costruiti utilizzando tali componenti mantengono tra loro una meta coerenza imposta dal sistema in *the large*, circostanza che può essere utilmente impiegata nella traduzione di un'opera ipermediale da un sistema ad un altro. Potremmo descrivere gli ipertesti costruiti secondo questo approccio usando la terminologia adottata in [Garz94] come segue: l'applicazione in *the large* e' costituita da una *collezione* di *nodi* che, oltre a poter essere percorsa, può distribuire copie di *nodi*, di *composite component* o slot ad un ipermedia cliente. Il prototipo presentato in questo lavoro esplora questa alternativa attraverso la costruzione di una collezione chiamata *G_lossario* ed e' attualmente limitato allo scambio tra opere ipertestuali prodotte con un medesimo ambiente di sviluppo.

Il lavoro e' organizzato come segue. Nella sezione uno presentiamo brevemente le caratteristiche degli oggetti *G_lossario* e *D_definizione*. Nella sezione due diamo conto di differenti approcci alla costruzione di dizionari e di differenti strategie di implementazione delle loro interfacce. Nella sezione tre descriviamo le modalità di interazione di un oggetto *G_lossario*. Nella sezione quattro,

infine, discutiamo alcuni aspetti implementativi del prototipo riguardanti l'algoritmo di lemmatizzazione adottato.

1 La classe di oggetti *G_lossario* e la classe *D_efinizioni*.

Il glossario realizzato nell'ambito di questo lavoro puo` essere considerato come appartenente ad una classe di oggetti denominata *G_lossario* (una collezione) e parzialmente composta di *nodi* che contengono *composite component* (oggetti) *D_efinizione*. Ogni occorrenza della classe *D_efinizione* rappresenta la definizione di un termine presente nel glossario. Un glossario puo` dunque essere visto idealmente come composto di un numero illimitato di *D\_efinizioni*. *G\_lossario* contiene anche nodi appartenenti alla classe *I\_ndex* e oggetti della classe *M\_enu`*. Gli oggetti appartenenti alla classe *I_ndex* permettono all'utente di accedere a differenti liste alfabetiche dei termini inseriti nel glossario, mentre gli oggetti appartenenti alla classe *M_enu`* permettono di pilotare le differenti opzioni disponibili all'interno dell'applicazione.

In figura 1 e` mostrata la composizione di un generico glossario secondo la tecnica Object Modelling Technique (OMT) [Rumb91].

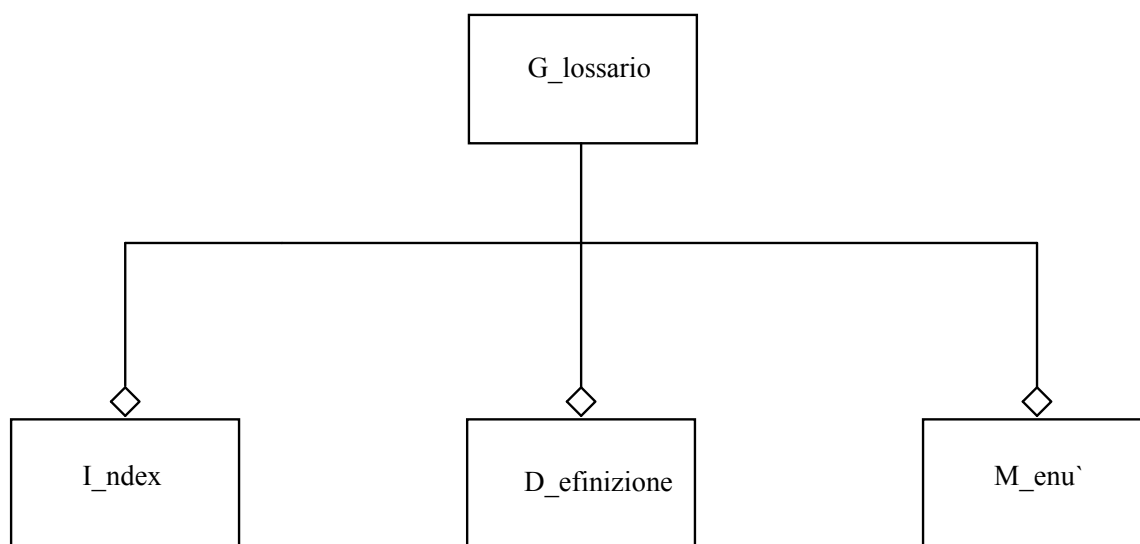


Figura 1: la classe *G_lossario*.

Un numero illimitato di glossari possono coesistere e piu` ipertesti cliente possono accedere ai dati in essi contenuti. Ogni glossario presentera` proprieta` comuni alla classe *G\_lossario* e differira` dagli altri solamente per il dominio semantico a cui appartengono le definizioni.

Gli utenti abilitati possono variare gli oggetti *D_efinizione* aggiungendo nuovi termini o cambiando il contenuto delle definizioni. Cio' e' possibile come spiegheremo nella sezione quattro, perche' gli oggetti della classe *G_lossario* adottano un algoritmo dinamico di lemmatizzazione, sicche' questi utenti possono interagire con un *G_lossario* mediante due modalita':

- possono accedere ad uno specifico *G_lossario* e creare, modificare o cancellare oggetti *D_efinizione*;
- possono eliminare, creare o sostituire oggetti *D_efinizione* agendo nel proprio ipertesto cliente.

Infine, utenti abilitati possono creare nuove occorrenze di oggetti *G_lossario*.

Coloro che non sono abilitati a modificare un oggetto *G_lossario* interagiscono anch'essi secondo una doppia modalita'.

Essi infatti possono consultare *G_lossario* o richiederne i servizi. Nel primo caso questi utenti navigano tra i nodi del glossario. Nel secondo caso ricevono, a seguito di una richiesta, una copia legale di un oggetto della classe *D_efinizione* che diventa parte dell'opera ipertestuale del lettore.

Il *G_lossario* include un algoritmo di *lemmatizzazione dinamica* che permette di ricondurre diverse occorrenze di un termine ad un unico lemma. Cosicche' tutti i diversi tempi e modi di un verbo per esempio amare (*amassi*, *amassimo*, etc.) verranno ricondotti al lemma *amare*. L'algoritmo permette la trattazione di tutte le parti del discorso regolari della lingua italiana ed inglese. Puo' essere adattato dagli utenti abilitati, i quali possono indicare la lista degli oggetti lessicali di base che vengono utilizzati dalla lemmatizzazione; tale modifica permette l'adattamento del glossario a lingue diverse dall'italiano o dall'inglese. Un'occorrenza della classe *G_lossario* puo' quindi contenere termini appartenenti a piu' lingue.

La classe *D_efinizione* e' composta da *composit components* che identificheremo d'ora innanzi come *foglietti*. I foglietti contengono il testo e la grafica che costituisce la definizione dei termini contenuti nel glossario. Un foglietto puo' contenere ogni tipo di dato: testo grafica o altri oggetti. Cio' significa che l'utente e' libero di creare ogni genere di definizione dei termini. Tutti gli oggetti elementari che possono essere utilizzati all'interno del sistema di sviluppo adottato per la costruzione del glossario possono essere utilizzati all'interno della definizione di un termine; ne consegue che le proprieta' degli oggetti elementari sono conservate quando viene eseguita una copia del foglietto. L'utente abilitato e' dunque lasciato libero di definire le caratteristiche dei *foglietti* e non e' costretto in nessun modo ad utilizzare schemi predefiniti. Questa liberta' compositiva dovra', probabilmente, essere limitata nello sviluppo dei prossimi prototipi, perche' essa discende direttamente dalla omogeneita' dell'ambiente di sviluppo. In un ambiente non omogeneo, in cui le richieste possono pervenire da ipertesti sviluppati con un'altro sistema, si richiede una traduzione del foglietto in formato accettato dal chiamante e questo, quasi sicuramente, comportera' una limitazione nella scelta degli oggetti elementari che costituiscono i foglietti.

Esistono comunque tre sub-classes nella classe *D_efinizione* (figura 2): testo, grafica e testo-grafica. I *foglietti* appartenenti a tali classi sono predefiniti in modo da facilitare l'inserimento di definizioni prevalentemente testuali, grafiche o miste.

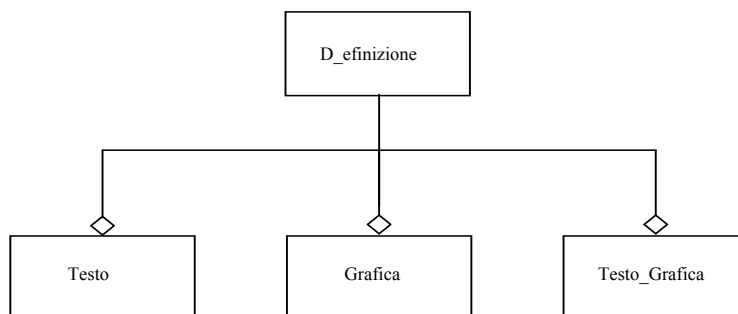


Figura 2: la classe *D_efinizione*.

Un'importante caratteristica della classe *D_efinizione* e' costituita dalla possibilita' di associare ai *foglietti* le informazioni relative al *copyright* delle definizioni. Cio' puo' avvenire relativamente al testo delle definizioni, alle immagini grafiche contenute nella definizione stessa o ad entrambe. Tali informazioni rimangono associate al foglietto anche nel caso in cui questo venga importato da altri ipertesti cliente, e nel caso in cui le informazioni vengano salvate in un file su richiesta dell'utente. Poiche' le copie dei foglietti sono oggetti standard, in un ipertesto cliente di un glossario possono essere implementati metodi che agiscono sui foglietti importati. Un esempio di funzione implementabile e' la creazione di un *G_lossario* speciale costruito con le copie provenienti da piu' *G_lossari*, oppure e' la funzione che consente ad un utente di catturare e salvare in un file ASCII qualsiasi testo sia presente in un *foglietto*, conservando automaticamente anche le informazioni relative al *Copyright* associato al testo stesso.

Come gia' accennato, numerosi ipertesti cliente possono consultare i dati contenuti nel glossario. Il glossario rispondera' alle richieste degli ipertesti abilitati inviando il *foglietto* corrispondente alla

definizione del termine richiesto. Una volta importati nell'ipertesto cliente, i *foglietti* mantengono tutte le proprietà presenti nel glossario, alle quali inoltre si aggiungono alcuni controlli prima non disponibili. Tali controlli permettono lo spostamento e la chiusura dei *foglietti* nell'applicazione cliente. Un numero illimitato di foglietti può essere importato in un generico ipertesto cliente da un numero illimitato di glossari appartenenti alla classe *G_lossario*. I *foglietti* importati, pur mantenendo le proprietà della classe *D_efinizione*, come ad esempio il *copyright* del termine, diventano parte integrante dell'ipertesto cliente.

Se ad un componente di un oggetto *D_efinizione* è associata l'ancora di partenza di un link, questa è conservata durante il processo di copia. Però nel primo esempio applicativo realizzato, un glossario di botanica, questa funzionalità non è stata volutamente adoperata. Il meccanismo è sostanzialmente analogo a ciò che avviene quando si copia localmente un documento WWW: le ancore definite nel documento sono conservate.

2. Dizionari e loro interfacce

Tradizionalmente, i dizionari macchina vengono raggruppati in diversi gruppi secondo diversi criteri. Una prima distinzione, dovuta a Raffaelli e Raspollini [Raff87], classifica i dizionari macchina in tre diversi gruppi.

- Un primo gruppo di dizionari macchina è rappresentato da semplici trascrizioni di dizionari originariamente presenti su supporto cartaceo. Le informazioni vengono spesso inserite senza una struttura evoluta, massimizzando la velocità di accesso ai dati e minimizzando l'occupazione di memoria. Sono esempi di questo tipo i dizionari utilizzati negli spell-checkers.
- Un secondo gruppo di dizionari macchina comprende i dizionari che sono basati su informazioni selezionate, classificate e codificate. In questo caso assume rilevanza non più una generica parola appartenente al dizionario (come nei dizionari del primo tipo) ma il *lemma*, che rappresenta l'entità *elementare* indirizzabile, con la corrispondente categoria grammaticale.
- Il terzo ed ultimo gruppo comprende i dizionari che vengono chiamati "basi di dati lessicali", in quanto forniscono una rappresentazione elaborabile dell'intera informazione lessicale.

Un'altra distinzione possibile è quella tra "*Machine - readable dictionaries*" (*MRD's*), ovvero dizionari disponibili in formato leggibile alla macchina e *Machine - tractable dictionaries* (*MTD's*), dizionari in formato *trattabile* da un processore e spesso appropriati per utilizzi nel campo del *natural language processing*. Un *MRD* è, in pratica, assimilabile ad un dizionario del primo e secondo gruppo della distinzione di Raffaelli e Raspollini, mentre un *MTD* è assimilabile ad uno del terzo gruppo.

L'utilizzo di glossari in ambienti interattivi e le strategie di lettura seguite dal lettore durante la consultazione di un testo, sono state oggetto di numerosi studi: [Guth87], [March89], [Blac91], [Wrig90], [Blac92]. In particolare Black A., Wright P., Black D. e Norman K. [Blac92] individuano quattro diverse modalità di interazione con un glossario all'interno di un ipertesto.

In primo luogo il glossario può, una volta che sia stato richiamato dall'utente tramite un "click" sul termine ignoto, comparire sovrapponendosi al testo precedentemente in lettura (figura 3). In questo caso, il lettore sarà costretto ad uscire completamente dal contesto di lettura in cui si trova.



Figura 3: sovrapposizione del glossario sull'ipertesto

La seconda possibilità è rappresentata dal fatto che il glossario può, una volta richiamato, occupare una sola parte dello schermo in modo da non obbligare l'utente ad uscire completamente dal contesto di lettura in cui si trova al momento della richiesta (figura 4).

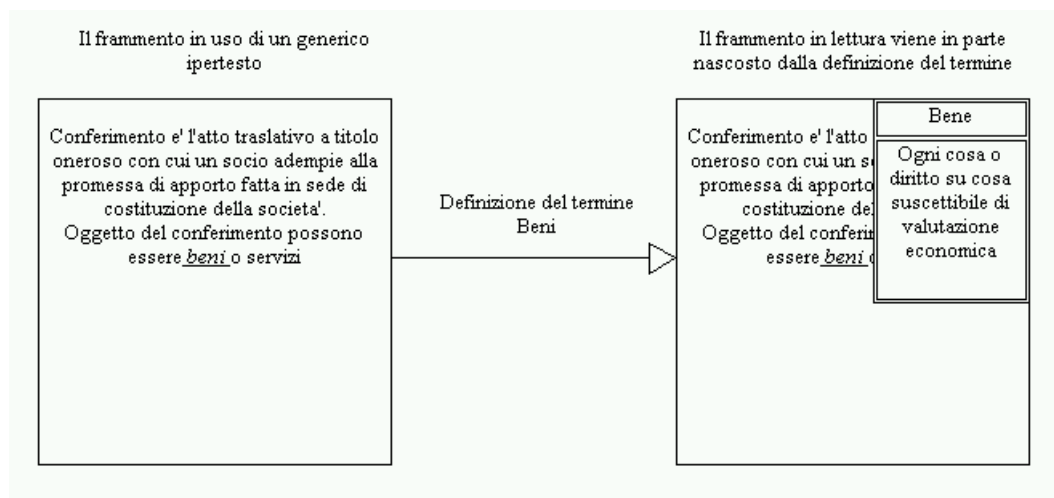


Figura 4: il glossario occupa una sola parte dello schermo

La terza modalità è rappresentata dal fatto che il glossario può essere visualizzato in una finestra sempre presente, di piccole dimensioni, che non impedisca il normale svolgimento della lettura del testo, dove, a richiesta, sia possibile far comparire un breve commento - spiegazione del termine selezionato (figura 5).

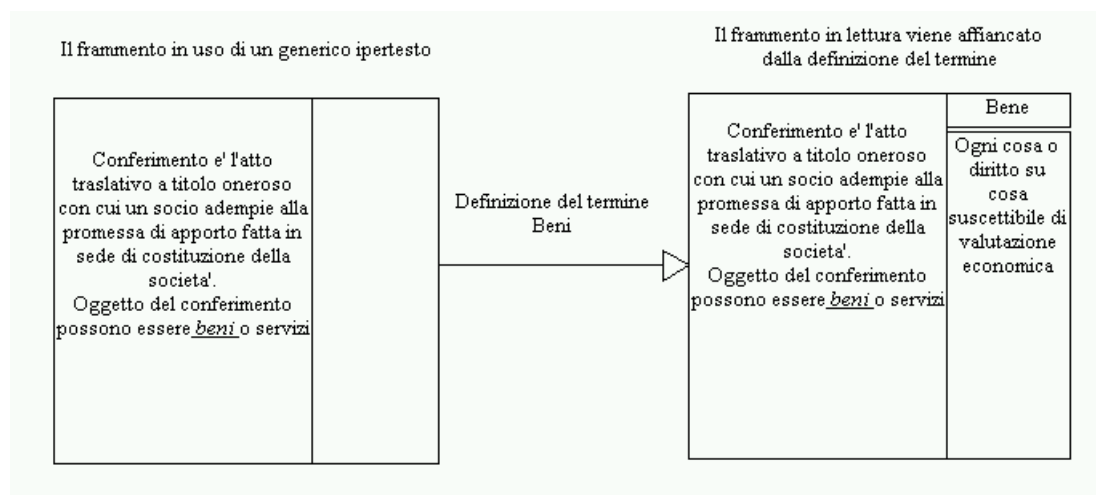


Figura 5: il glossario è contenuto in una finestra separata dall'ipertesto

La quarta ed ultima modalita' di visualizzazione, implica il fatto che il glossario possa essere richiamato dall'utente tramite un "click" sul termine ignoto. I termini richiamabili sono permanentemente visibili in una lista - glossario affiancata all'ipertesto. Questa lista, contiene i termini presenti nel glossario per quel peculiare contesto di lettura. (figura 6).

Frammento di ipertesto in lettura	Termini disponibili nel glossario
<u>Conferimento</u> e' l'atto <u>traslativo</u> a titolo oneroso con cui il socio adempie alla promessa di <u>apporto</u> fatta in sede di costitu- zione della <u>societa'</u>. Oggetto del conferimento possono essere <u>beni</u> o <u>servizi</u>	apporto bene conferimento servizio societa' traslativo

Figura 6: i termini presenti nel glossario sono visibili in una lista affiancata all'ipertesto.

Marchionini [March89], invece, si occupa di verificare come si modifichi il comportamento dei lettori nel passaggio da una enciclopedia di tipo tradizionale ed una di tipo elettronico con funzionalita' ipertestuali. Piu' precisamente, Marchionini analizza come i "*modelli mentali*" utilizzati normalmente dal lettore per la consultazione di una enciclopedia tradizionale vengono impiegati e come si modificano passando alla versione elettronica della stessa. Marchionini, inoltre, si occupa di analizzare sotto quali condizioni l'introduzione della versione elettronica dell'enciclopedia comporti un miglioramento nelle possibilita' offerte al lettore dell'opera. Le conclusioni di Marchionini, compiuta utilizzando le versione scritta e quella elettronica della *Grolier's Academic American Encyclopaedia* possono essere riassunte in due punti:

- l'utente dell'inclopedia tradizionale dispone di "*modelli mentali*" sufficientemente evoluti per la proficua consultazione dell'enciclopedia elettronica;
- il passaggio dalla versione tradizionale alla versione elettronica ha fornito risultati sufficienti ma tuttavia non ottimali, in quanto gli utenti, adattando i "*modelli mentali*" normalmente utilizzati per la versione tradizionale dell'enciclopedia, hanno ignorato molte modalita' di ricerca disponibili nella versione elettronica, utilizzando la stessa in modo equivalente a quanto avrebbero fatto per la versione cartacea.

Secondo Marchionini, dunque, le realizzazioni di enciclopedie in formato elettronico devono essere in grado di permettere l'adattamento degli utenti abituati all'uso di versioni cartacee, facendo in modo che l'utente possa trovare somiglianze tra la versione cartacea e quella elettronica, adattando progressivamente i propri modelli mentali alle nuove potenzialita' offerte.

Il glossario qui presentato non segue perfettamente nessuno dei quattro approcci indicati da Black A., Wright P., Black D. e Norman K. [Blac92]. La modalita' uno puo' infatti essere ottenuta consultando *G_lossario* direttamente; in questo caso il nodo che contiene l'oggetto *D_efinizione* puo' sovrapporsi alla finestra su cui compare il testo in lettura. Le modalita' due, tre e quattro (comunque ottenibili) impongono limiti compositivi nell'impaginazione di un ipermedia, perche' richiedono tutte che una porzione dello schermo sia dedicata alla interazione con il glossario. Inoltre in tutti e tre i casi gli oggetti che permettono la visualizzazione delle definizioni permettono la visualizzazione di una definizione alla volta. Nella soluzione quattro infine e' presupposta una prenoscenza dei requisiti del lettore. Nell'esempio di figura 6 il termine *societa'* potrebbe non essere indicato nella lista di destra, se si presuppone che il lettore sia relativamente specializzato,

mentre potrebbe essere essenziale se si presuppone che il lettore sia digiuno di ogni terminologia giuridica.

In primo luogo, ogni termine che compare in un testo può essere oggetto di ricerca; la scelta dipende dalla maggiore o minore familiarità del lettore con il lessico adoperato. Inoltre la funzione svolta da una definizione è prevalentemente temporanea, una volta compreso il significato di un termine, non è più necessario disporre di una definizione. La soluzione delle copie permette una estrema libertà: Si possono richiedere copie di più definizioni, si possono conservare le copie, si possono iconizzare le copie e quando non servono più si possono tranquillamente distruggere. Infine, nel nostro approccio, ogni foglietto è dotato di metodi che ne permettono lo spostamento sullo schermo, il che permette di annullare i vincoli compositivi derivanti da partizioni fisse dello schermo.

Queste funzionalità sono permesse dalle particolari caratteristiche della classe *D_efinizioni*. Il glossario, tuttavia, segue i suggerimenti indicati da Marchionini [March89]. Infatti, come già visto nella sezione uno di questo lavoro, il glossario può essere utilizzato secondo due modalità distinte. Nella prima può essere utilizzato come applicazione *stand-alone*, mentre nel secondo caso i dati presenti nel glossario possono essere consultati per mezzo di un qualsiasi ipertesto cliente abilitato. Nel primo caso, l'utente proveniente da versioni cartacee di glossari potrà usare le funzionalità basilari del programma che gli permetteranno di utilizzare i "modelli mentali" già posseduti. Successivamente potrà utilizzare le funzioni di ricerca avanzate e le opzioni che permettono di richiamare le definizioni attraverso un ipertesto cliente.

3 Consultazione di un G_lossario e colloquio con un G_lossario

Come abbiamo anticipato nella premessa, l'obiettivo del lavoro è quello di sperimentare ambienti didattici basati sulla tecnologia degli ipertesti. In questa prospettiva abbiamo bisogno di un approccio che garantisca una consistenza semantica agli ipertesti prodotti senza trasformare gli stessi in una versione sofisticata di un sistema di istruzione programmata. Un utente che crei una occorrenza di un *G_lossario* ne acquista i diritti di autore che può eventualmente estendere ad altri. Un utente che abbia il diritto di autore su un *G_lossario* può: aggiungere, modificare e cancellare gli oggetti *D_efinizione* di uno specifico glossario sia interagendo con *G_lossario* direttamente, sia interagendo da un ipertesto cliente di *G_lossario*. Al contrario, un utente che abbia solo diritti di lettura non può eseguire queste azioni, ma solo quelle di consultazione del *G_lossario* e di richiesta di copie di oggetti *D_efinizione*.

La consultazione del glossario può essere effettuata in quattro modi diversi:

- indicando specificatamente il termine da ricercare in una dialog box richiamata attraverso l'apposita opzione presente in uno dei menu appartenenti alla classe *M_enu*;
- evidenziando un termine all'interno di un qualsiasi testo del glossario e premendo il tasto di destra del mouse;
- scegliendo il termine dalle liste alfabetiche, che compongono l'oggetto *I_ndex*;
- seguendo un *Guided_Tour* [Zell89] che permette di passare da un nodo al successivo o al precedente in ordine alfabetico in un *G_lossario*.

I quattro tipi di ricerca sono in parte differenti, infatti mentre la ricerca secondo le modalità 1 e 2 avviene attraverso il meccanismo di lemmatizzazione, il quale può produrre anche una risposta negativa qualora il glossario non contenga il lemma cercato, la ricerca nell'indice alfabetico avviene direttamente sui soli lemmi presenti nel glossario. Infatti, se una definizione è presente nel glossario, allora esiste un nodo (ovvero unità minima di destinazione di un link) che ha come nome quel lemma. Un algoritmo di indicizzazione aggiorna automaticamente le liste alfabetiche ad ogni inserimento di un nuovo lemma e le mantiene ordinate. Le liste alfabetiche di un oggetto *I_ndex* sono attive. Selezionare in esse un lemma ed eseguire un click col tasto sinistro del mouse significa eseguire un salto al nodo il cui nome coincide con il lemma selezionato. Le stesse liste sono usate per gestire il path del *Guided Tour* alfabetico.

Nel caso invece che la richiesta avvenga mediante l'indicazione di una stringa, oppure mediante la selezione di un termine in un testo del glossario stesso, il termine indicato viene sottoposto all'algoritmo di lemmatizzazione per estrarne il lemma che diviene l'argomento della funzione di retrieval.

Questo modo di operare permette ad un utente che sta consultando una *D_efinizione* di *G_lossario* di utilizzare come argomento di ricerca una porzione di testo che compare nel testo della *D_efinizione*. Questo modo di ricercare si presta ad eseguire ricerche concatenate potenzialmente illimitate perché ogni definizione di un glossario può contenere termini non noti al lettore. Consideriamo, ad esempio, un glossario di termini giuridici. L'utente, dopo aver richiamato la definizione del termine "società" ad egli sconosciuto nella sua accezione giuridica, potrà richiamare la definizione del termine "conferimenti" presente nella definizione del termine "società" e così via. Casi in cui questo tipo di ricerche sono necessarie non sono rari tra gli utenti di un glossario.

Ogni generico ipertesto può diventare cliente di un oggetto appartenente alla classe *G_lossario*. A tal fine, sarà sufficiente che esso invii una specifica richiesta al *G_lossario*, il quale restituirà i metodi necessari all'ipertesto cliente per la sua abilitazione. Tali metodi riguardano l'aggiunta di un menu all'interno dell'applicazione cliente, per l'interazione con un *G_lossario* e le primitive per la comunicazione. I metodi passati sono nella forma sorgente e quindi eventualmente modificabili dall'autore o dal lettore dell'ipertesto cliente. La trasmissione dei metodi avviene attraverso l'invio di un composit component che ha l'aspetto di una icona. Una volta attivato, attraverso un click con il tasto sinistro del mouse, l'oggetto installa i metodi nell'ipertesto cliente e si autodistrugge.

Le modalità di interazione disponibili all'interno di un'applicazione cliente rispecchiano quelle presenti all'interno del glossario stesso e sono:

- l'apertura del *G_lossario* in modalità di consultazione;
- indicando specificatamente il termine da ricercare in una dialog box richiamata attraverso l'apposita opzione presente in uno dei menu;
- evidenziando un termine all'interno di un qualsiasi testo presente nell'ipertesto e premendo il tasto di destra del mouse.

Anche quest'ultima modalità presenta un funzionamento simile a quella presente nel glossario stesso. La differenza principale è che in questo caso, invece di eseguire un salto al nodo del glossario che contiene la definizione del termine scelto, comparirà un foglietto proveniente dal glossario e contenente la definizione del termine richiesto. Questo *foglietto* presenterà tuttavia funzionalità normalmente non disponibili all'interno del glossario. Infatti all'atto dell'operazione di copia all'oggetto definizione vengono aggiunti i metodi che permettono di cancellare la copia dopo averla utilizzata, iconizzare il foglietto nell'ipertesto chiamante e infine spostare il foglietto nel nodo di arrivo.

La possibilità offerta agli utenti-autori di aggiungere nuovi oggetti *D_efinizione* in modo dinamico è facilitata dal particolare algoritmo di *lemmatizzazione* utilizzato. L'algoritmo differisce in modo sostanziale dagli algoritmi *statici* normalmente utilizzati nella costruzione di glossari e dizionari, i quali, generalmente non permettono la trattazione di nuovi termini da inserire nel glossario. I nuovi termini in questi glossari vengono semplicemente mantenuti distinti da quelli inseriti all'origine. In questo caso, dunque, i nuovi termini verranno semplicemente esclusi dalla fase di lemmatizzazione e la possibilità di aggiornamento del glossario da parte dell'utente risentirà in modo sostanziale di questa limitazione. Al contrario, l'algoritmo qui utilizzato è in grado di trattare i nuovi termini in modo equivalente ai termini già presenti, non escludendoli dalla fase di lemmatizzazione e senza la necessità di disporre di un lessico di base. I nuovi *foglietti* godranno di tutte le proprietà della classe *D_efinizione* e saranno trattati dall'oggetto *G_lossario* in modo esattamente eguale ai termini inseriti precedentemente o all'origine. Ciò permette una elevata flessibilità di utilizzo.

4. Note implementative

In questa sezione viene discussa in dettaglio l'implementazione del glossario e gli algoritmi utilizzati. Il glossario qui descritto deve essere inteso come un generico oggetto glossario appartenente alla classe *G_lossario*.

Il glossario e' scritto parte in *OpenscriptTM* e parte in C, relativamente all'algoritmo di lemmatizzazione compilato in una *dynamic link library* di *WindowsTM*. Il programma utilizza anche la libreria standard di *ToolbookTM* che permette la gestione dei files in formato DB3. Gli ipertesti cliente abilitati all'accesso ai dati contenuti nel glossario includono un modulo scritto in *OpenscriptTM* per il dialogo e lo scambio di dati con il glossario. Il sistema e' installabile su di un qualsiasi sistema MS-dos dotato di Microsoft Windows e di Multimedia Toolbook 3.0. Tuttavia, l'ambiente consigliato consiste in un PC basato su 486 66Mhz con 8Mbytes di memoria e scheda grafica 1024*768 in 64.000 colori.

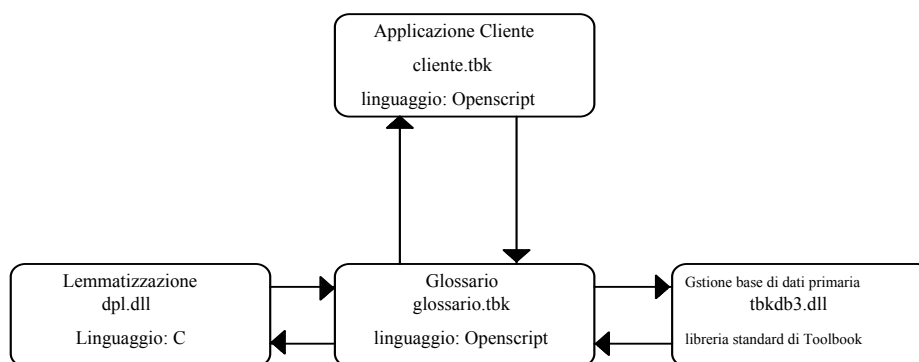


Figura 7: componenti principali del glossario.

Il glossario risponde dunque alle chiamate provenienti dall'interfaccia utente residente all'interno del glossario stesso o ad eventuali richieste provenienti da ipertesti cliente abilitati. Più di un oggetto glossario può essere attivo contemporaneamente in memoria. Ciò significa che gli oggetti glossario dovranno essere in grado di individuare l'ipertesto chiamante ed indirizzare ad esso le appropriate risposte, consistenti in messaggi di successo, insuccesso e eventuali *foglietti* contenenti le definizioni dei termini richiesti.

4.1 L'algoritmo di lemmatizzazione.

L'algoritmo di *lemmatizzazione* e' scritto in linguaggio C e risiede in una *dynamic link library* esterna al glossario stesso. L'algoritmo utilizzato e' di tipo *dinamico*. Ciò significa principalmente che l'algoritmo e' in grado di agire autonomamente, senza l'intervento dell'utente del dizionario o la disponibilita' di un lessico di supporto per l'ottenimento del lemma del termine. L'algoritmo di lemmatizzazione e' inoltre di tipo *morfologico*, in quanto i lemmi individuati dall'algoritmo non sono necessariamente grammaticalmente corretti. L'algoritmo di lemmatizzazione agisce sia nella fase di ricerca di termini già presenti nel glossario, sia nell'inserimento di nuovi termini, avvertendo l'utente qualora un termine che presenti lo stesso lemma sia già presente nel glossario. L'algoritmo di lemmatizzazione, essendo di tipo *dinamico*, non e' *esaustivo* ne' tantomeno *non-generativo*. Un algoritmo di tipo *statico* simile a quello utilizzato, ad esempio in [Raff89], avrebbe permesso prestazioni migliori, ma avrebbe comportato l'impossibilita', per l'utente, di poter aggiornare i dati del glossario se non escludendoli dalla fase di lemmatizzazione stessa e avrebbe impedito la possibilita' di personalizzare il glossario per l'utilizzo con lingue diverse dall'italiano. Secondo il nostro approccio l'utente deve poter essere in grado di creare nuovi oggetti *G_lossario* ed utilizzarli per inserire un'arbitraria scelta di termini, quindi un algoritmo di lemmatizzazione di tipo *statico* non sarebbe utile. D'altra parte, l'utilizzo di tecniche dedicate come, ad esempio, la tecnica degli *automi a stati finiti* [Luch93] non e' stata ritenuta adeguata, in quanto la gestione di particolari funzioni, specificatamente quelle relative alla gestione della base di dati, ad esempio la

produzione di una lista alfabetica dei termini inseriti nel glossario, sarebbe risultata inutilmente complessa.

L'algoritmo è in grado di identificare correttamente i lemmi di tutte le parti del discorso regolari della lingua italiana: sostantivi, preposizioni semplici, articoli, preposizioni articolate, avverbi, pronomi (personali, relativi, possessivi, dimostrativi, indefiniti, interrogativi, esclamativi, numerali), aggettivi, verbi, congiunzioni. L'algoritmo prevede tre passi successivi per la ricerca di un lemma:

- a) Il termine indicato dall'utente viene ricercato nel dizionario. Qualora non venga trovato, si verifica la lunghezza del termine da ricercare. Qualora essa sia minore o eguale a due, si suppone che il termine rappresenti un articolo o una preposizione semplice, che appartenerebbero comunque a lemmi diversi e dunque la ricerca termina alla fase a); se invece il termine è di lunghezza maggiore di due, si può supporre che rappresenti un'altra parte del discorso (sostantivi ecc.) e dunque vengono eseguite la fase b) ed eventualmente la fase c).
- b) In questa fase, dal termine non trovato della fase a), viene eliminata l'ultima lettera, ottenendo l'ipotetico lemma e viene ripetuta la ricerca. Questo permette di considerare tutte le possibili variazioni delle seguenti parti del discorso: nomi, preposizioni articolate, aggettivi (modo positivo e comparativo), pronomi (per quanto attiene a quelli declinabili).
- c) Qualora il termine non sia ancora stato trovato, si può ritenere che, se esiste, debba considerarsi o un verbo, oppure un aggettivo superlativo. Viene dunque eseguita la fase di lemmatizzazione vera e propria: si confronta il termine con tutti i possibili suffissi dei verbi regolari (*are, ere, ire, ecc.*) e degli aggettivi superlativi. Qualora venga verificato che la desinenza coincide con uno dei suffissi conosciuti, questo viene eliminato dal termine stesso, ottenendo quello che si suppone essere il lemma morfologico del termine. La ricerca del termine viene dunque ripetuta una terza volta nel dizionario e, qualora il lemma non corrisponda a nessun termine memorizzato, la ricerca fallisce definitivamente.

L'algoritmo effettua un numero di ricerche pari a tre (per verbi e aggettivi superlativi) e pari a due nel maggior numero dei casi (sostantivi). La fase di lemmatizzazione è liberamente personalizzabile dall'utente, il quale può modificare i suffissi utilizzati dall'algoritmo.

Ogni glossario appartenente alla classe *G_lossario* dispone di due basi di dati distinte:

- una base di dati primaria ad accesso veloce dove viene memorizzata la lista dei termini disponibili nel glossario e viene gestita per mezzo di files indicizzati memorizzati in formato DB3;
- la base di dati secondaria, mantenuta all'interno del glossario stesso, dove vengono memorizzati, distinti per ogni termine, i dati relativi al testo delle definizioni, le immagini ed eventuali oggetti ipertestuali.

La distinzione tra le due basi di dati è dovuta principalmente alla relativa lentezza di *Toolbook* nel permettere l'accesso ai dati mantenuti internamente ad una sua collezione. Di conseguenza, la base di dati primaria agisce da filtro alla base di dati secondaria che viene interrogata solo qualora l'operazione richiesta (inserimento, ricerca, modifica etc.) abbia avuto successo nella base di dati primaria.

Conclusioni

Il prototipo di glossario che abbiamo presentato può essere considerato un esempio di ciò che intendiamo per materiali didattici ipertestuali, ovvero applicazioni costruite come oggetti generali che rispettano i paradigmi delle tecnologie ipertestuali e che possono essere utilizzate come strumenti nel corso di un processo di apprendimento. L'uso delle tecnologie degli ipertesti nei processi di apprendimento infatti non riguarda soltanto la possibilità di navigare tra una collezione di pagine come avviene spesso negli ipertesti distribuiti sul mercato. Se il contributo apportabile da queste tecnologie fosse limitato al retrieval di notizie, i processi di apprendimento non ne

verrebbero particolarmente favoriti. Alcuni potrebbero sostenere, non senza qualche ragione, che, poiche' ad ogni processo di apprendimento reale e' associata una certa quantita' di lavoro, cio' che si trova rapidamente si dimentica anche molto rapidamente e quindi tutto cio' che riduce il lavoro va nella direzione opposta allo sviluppo dell'apprendimento. Noi riteniamo che queste tecnologie mostrino tutto il loro potenziale se usate come strumento per lo sviluppo di autonome capacita' intellettuali da parte dei soggetti del processo di apprendimento.

Nel lavoro abbiamo presentato il prototipo di un glossario utilizzabile come strumento in ambienti didattici che utilizzino tecnologie ipermediali. Abbiamo descritto nella sezione uno la natura degli oggetti che compongono il dizionario, nella sezione due abbiamo confrontato l'approccio seguito con altre realizzazioni dello stesso tipo, mentre nella sezione tre abbiamo descritto le modalita' di interazione degli utenti con il glossario. Infine nella sezione quattro abbiamo sommariamente spiegato l'algoritmo di lemmatizzazione utilizzato.

Bibliografia:

- [Blac91] On-line consultation of definitions and examples: implications for the design of interactive dictionaries, in *Applied Cognitive Psychology*, pag. 149--166, No. 5, 1991.
- [Blac92] Black A., Wright P., Black D., Norman K., Consulting on-line Dictionary Information While Reading, in *Hypermedia*, pag. 145--169, Vol 4, No.3, 1992
- [Cola95] Colazzo L., Molinari A.: To see or not to see:tools for teaching, with hypertext slides, in printing in *EDMEDIA 95*, Graz giugno 1995
- [Garz94] Garzotto F.,Mainenti L.,Paolini P.: Adding Multimedia Collections to the Dexter Model, in *Proc.ECHT 94*, Edinburg, 18-23 September, 1994- 70-80
- [Guth92] Guthrie L., Guthrie J., Wilks Y., Cowie J., Farwell D., Slator B., Bruce R., A research program on machine-tractable dictionaries and their application to text analysis, New Mexico State University, 1992.
- [Legg94] Legget J.J.,Schnase J.J:Viewing Dexter whit open Eyes, in *Communications of the ACM*,Vol. 37, N 2, February 1994, 76-86
- [Luch93] Lucchesi C.K., Kowaltowski T., Applications of Finite Automata Representing Large Vocabularies, in *Software-Practice and Experience*, pag. 15/30, Vol. 23, January 1993.
- [March89] Marchionini G., Making the transition from print to electronic encyclopaedias: adaption of mental models, in *International Journal of Man-Machine Studies*, pag. 591--618, No.30, 1989
- [Raff87] Raffaelli R, Raspollini C., Il sistema ABCD (A Basic Computer Dictionary): struttura e realizzazione di un prototipo di dizionario della lingua italiana, IBM Italia - Centro studi di Pisa, 1987.
- [Rumb91]Rumbaugh J., Blaha M., Premerlani W., Eddy F., Lorensen W., Object-Oriented Modeling and Design, General Electric Research and Development Center Schenectady, New Yoark, 1991, Prentice-Hall International.
- [Wrig90] Wright P., Hull, Black A., Black D., Integrating diagrams and text, in *The Technical Writing Teacher*, pag. 244--254, N.17, 1990
- [Zell89] Zelleweger P.:Scipted documents: A hypermedia Path mechanism., in *proc. ACM Hypertext 89*, Pittsburgh, PN, nov.1989.