APPENDICE:

SCHEDE TECNICHE
DOCUMENTI

SCHEDA 1: DATABASE O BASI DI DATI

La seconda applicazione in ordine di importanza e utilizzo, dopo la videoscrittura, è la gestione – intesa come memorizzazione, elaborazione e utilizzazione – delle informazioni più strutturate di cui disponiamo.

Sino a qualche anno fa si era soliti parlare di «banche dati», ponendo l'accento sull'aspetto conservativo e di *information retrivial* (ossia reperimento delle informazioni) proprio di queste applicazioni, oggi si preferisce la denominazione «basi di dati» che, invece, mette in evidenza l'aspetto fondante dell'informazione per qualsiasi operazione e ne afferma l'aspetto dinamico e relazionale.

Dal punto di vista dei «pacchetti applicativi» per la gestione di basi di dati, l'informazione è un «dato» che viene memorizzato in una forma precisa, in un posto determinato e all'interno di una struttura predefinita, in grado di renderne il ritrovamento e l'elaborazione il più agevole possibile.

La struttura centrale dei programmi di *data base* (chiamati anche *dbase*) è il *record* – paragonabile ad una scheda d'archivio– suddiviso nei diversi *field*, letteralmente campi, che contengono distinti fra loro gli elementi contenuti nei vari *record*. Ad esempio, una base di dati bibliografica sarà costituita di *record* che contengono informazioni sui vari volumi e ogni *record* sarà composto di un «campo-autore», di un «campo-titolo», di un «campo-data», di un «campo-editore» e di un «campo-luogo d'edizione».

SCHEDA 2: DTD O DOCUMENT TYPE DEFINITION

La struttura astratta di un testo – quella di cui si occupa lo standard Sgml della Tei – è descritta dichiarando in una speciale tabella, detta appunto *Document Type Definition*, ossia Dtd, gli elementi che la costituiscono e le relazioni che intercorrono fra questi elementi. Il linguaggio non indica alcuna prescrizione riguardo alla tipologia e al numero dei *tag*, ma indica solo le norme sintattiche che regolano la definizione dei marcatori all'interno della Dtd.

Lo Sgml ci permette di crearci da soli la migliore grammatica possibile per descrivere un determinato documento, ma la difficoltà dell'operazione ed una maggiore economia, tanto dei costi quanto dei tempi, consiglia di rivolgersi a una Dtd preesistente, ad esempio uno fra i diversi potenti strumenti forniti dalla Tei. Nella Dtd vengono dichiarati:

- I marcatori per gli «elementi» presenti nel testo (possono essere titolo, paragrafo, nota, pagina ecc.);
- La descrizione del «contenuto» di questi elementi e l'ordine con il quale elementi e contenuti possono comparire (un paragrafo potrà contenere parole e le parole potranno contenere lettere, il numero di pagina i numeri ecc.);
- Gli *attributi* da assegnare a ogni elemento (un titolo potrà essere contrassegnato dalla propria resa grafica, una nota dalla sua posizione nella pagina);
- I marcatori per le «entità».

Codificare un testo in Sgml/Tei consiste, quindi, non solo nell'inserire i *tag* che lo descrivono, ma significa trasformarlo in un oggetto costituito da tre parti:

- la *Sgml declaration*. Essa definisce quali caratteri sono usati nella Dtd e nel documento testuale, inoltre, definisce ogni tratto di Sgml usato nel documento e può essere memorizzato esternamente al documento. In caso di mancata dichiarazione ne è disponibile una di *default*;
- il *Document Type Definition*, che definisce le strutture e le regole da utilizzare per codificare il documento e contiene i caratteri definiti nella *Sgml declaration*. Anche la Dtd può essere conservata separatamente dal documento.
- la *Document Instance*. Il documento contiene in se stesso il testo e il riferimento alla Dtd, i suoi caratteri sono definiti nella *Sgml Declaration* ed è codificato in accordo con le regole definite nella Dtd.

SCHEDA 3: *E-BOOK*

L'espressione «libro elettronico», *e-book*, non corrisponde a un'unica definizione, ma in questa scheda eviteremo di entrare nel dibattito, proponendo una spiegazione il più possibile generale e imparziale.

In generale si può definire libro elettronico qualunque:

> testo compiuto, organico e sufficientemente lungo ('monografia'), eventualmente accompagnato da metadati descrittivi, disponibile in un qualsiasi formato elettronico che ne consenta – fra l'altro – la distribuzione in rete, e la lettura attraverso un qualche tipo di dispositivo hardware, dedicato o no. [1]

L'e-book nasce dall'evoluzione del libro tradizionale su carta, come conseguenza dell'informatizzazione dell'editoria. Infatti le nuove tecnologie hanno modificato profondamente i processi di stampa, rendendoli completamente digitali. L'effetto immediato a questa rivoluzione è stata l'esigenza di abbandonare il libro in carta.

Ma se lo studio di lettori per *e-book* e computer palmari quali e-paper e dell'*e-ink* sta avendo una grande evoluzione, in realtà questo settore sta facendo da pochi anni i suoi primi passi cercando di sfruttare al meglio le sue potenzialità, per permettere la lettura di testi elettronici in maniera più comoda di quella in ambiente elettronico.

Comunque, se si considerano libri elettronici anche quelli in cui il testo elettronico è unicamente il «supporto di trasferimento dell'informazione destinata alla stampa su carta» – come per i *print on demand* – essi sono uno strumento di distribuzione più comodo ed economico rispetto al libro a stampa, anche se questo rimane lo strumento più fruibile per la lettura da parte dell'utente finale.

[1] Gino Roncaglia, *Merzweb*, <http://www.merzweb.com>.

L'editoria digitale si trova ancora nell'ambito della sperimentazione: tenendo presenti le caratteristiche dei dispositivi hardware di lettura e la scomodità evidente del normale monitor da scrivania, si comprende che il mercato e-book non è destinato ad un boom di diffusione o di vendite immediato e spettacolare. Il libro elettronico ha il potenziale per sostituire, in alcune situazioni, il libro su carta, ma seguendo un lento processo, e dopo la progettazione di soluzioni hardware, software e commerciali migliori di quelle oggi sul mercato.

SCHEDA 4: HTML

Il Linguaggio *Html*, sigla per *Hyper Text Mark Up Language*, è un sottoinsieme di Sgml [2], una sorta di sua applicazione.

Html è il linguaggio alla base del fenomeno Web; è una Dtd creata dalla Tei [3] per creare una codifica che preparasse i documenti al viaggio attraverso la rete e possiede una struttura semplificata rispetto a quella dello Sgml.

In questo caso, infatti, lo scopo non è quello scientifico di conservazione dell'informazione, ma quello pratico della realizzazione di un «modello» del testo che garantisca la massima trasmissibilità e godibilità.

La caratteristica fondamentale dello *Html* rispetto allo *Sgml* è di concentrarsi sull'aspetto tipografico del documento e non sulla sua struttura. Se in *Sgml* di fronte a una parola in corsivo dovevamo chiederci cosa rappresentasse quel corsivo, quale fossa il suo significato (è un'enfasi? una parola straniera? un titolo?) per poterlo codificare correttamente e non perdere informazione, in *Html* ci limiteremo a segnalare la presenza del corsivo.

[2] Vedi Scheda dedicata allo Sgml e Dtd.
[3] Tei, *Text Encoding Intiative*, www.tei-c.org.

Facciamo un esempio.
<I>Html</I>

In questa riga possiamo riconoscere i caratteristici segnali di apertura echiusura di marcatori proprio di *Sgml* (<, >, </), intuiamo che la i maiuscola (I) sta per *italic*, cioè corsivo e vediamo che «Html» è la parola da rendere in corsivo, ma non si può fare a meno di notare che se nello stesso testo trovassimo

<I>L'isola del giorno prima</I> e <I>boat people</I>
non avremmo alcuna possibilità di distinguere le diverse funzioni svolte dai corsivi nel testo, partendo dalla codifica effettuata.

Html si appoggia alla semplice constatazione che gli elementi sicuramente riconosciuti da tutti i computer, ossia i primi 128 caratteri del codice Ascii [4], sono gli stessi che viaggiano senza fatica sulla rete, per fruttare le caratteristiche di codifica di un linguaggio di marcatura generico ai fini della trasmissione.

[4] Codice Ascii: *American Standard Code for Information Interchange. Standard* approvato dall'Ansi (*American National Standard Institution*) alla fine degli anni Sessanta. E' uno dei codici più usati per la rappresentazione dei caratteri alfanumerici.

SCHEDA 5: LIT E PDF

I formati Lit e Pdf sono due formati proprietari per la conservazione ed il «trasporto» di file testuali, cioè sono due standard per e-book creati da due case di produzione software private e, come tali, il loro codice sorgente rimane segreto ed il loro utilizzo rientra sotto le leggi del copyright.

Il formato Lit, sviluppato da quella che attualmente è la più grande e importante *softwarehouse* del mondo, la Microsoft di Bill Gates, è una versione compilata dello standard pubblico OEB (*Open eBook*) al cui sviluppo la Microsoft stessa ha fornito importanti contributi. La partecipazione di Microsoft alla creazione dello standard pubblico può essere considerata un tentativo di scalzare lo standard di fatto imposto negli anni dalla Adobe con il suo Pdf.

Nelle intenzioni dell"Open eBook Forum [5] il formato OEB dovrà costituire la 'lingua franca' del mondo e-book: l'adozione di tale formato avrebbe, infatti, il vantaggio di permettere, in prospettiva, la «compilazione» diretta del libro elettronico per un gran numero di piattaforme hardware e software, partendo da uno stesso «pacchetto» di file OEB. Al momento, il più diffuso di tali formati è quello utilizzato dal programma Microsoft Reader, il Lit appunto.

[5] Sito ufficiale dell'Open e-Book Forum all'indirizzo: <www.openebook.org>

Il formato Pdf, ossia Adobe® Portable Document Format, è lo standard de facto per la distribuzione di documenti diffuso oramai in tutto il mondo, prodotto dalla *softwarehouse* Adobe. Il formato PDF è un formato di file universale che preserva tutte i font, la formattazione, i colori e le immagini di qualsiasi documento sorgente, indipendentemente dall'applicazione e dalla piattaforma usate per crearlo. Nato come formato per il *delivery* (trasporto) di documenti in ambiente editoriale, si è imposto per la facilità con cui il suo utilizzo può essere esteso al mondo web ed anche per l'intelligente politica di distribuzione gratuita del suo software lettore.

I file PDF sono compatti e possono essere condivisi, visualizzati, consultati e stampati da chiunque grazie ad *Adobe Reader*, software che supporta le più diverse piattaforme hardware e software: dai pc IBM compatibili, ai Macintosh, ai server Unix.

Fino ad oggi, pecca del formato Pdf è la sua origine come formato vettoriale nato per «descrivere» l'immagine grafica della pagina e non tanto il contenuto, ma successive versioni e l'accoppiamento recente con il *Glassbook Reader*, lo rendono più flessibile.

SCHEDA 6: SGML

Lo *Standard Generalized Mark-Up Language* [6], oltre ad essere il più noto fra i diversi linguaggi dichiarativi [7], è quello che ha maggiore influenza nel settore umanistico, soprattutto grazie alle *Document Type Definition* – comunemente ricordate con la sigla Dtd– prodotte dalla Tei [8], che si propongono come punto di riferimento per chiunque si occupi di trattamento informatico dei testi, quindi anche per lo storico comunicatore del web.

Questo linguaggio si basa su un sistema di *mark-up*, in italiano marcatura, generico che punta a descrivere la struttura astratta di un documento piuttosto che il suo aspetto grafico e, a questo scopo, consente la dichiarazione di un insieme di marcatori, tecnicamente *tag*. Un *tag* è un elemento che definisce un altro elemento. Infatti, nell'operazione di codifica una «marca», o *tag,* è una parola del linguaggio di codifica che descrive determinate proprietà dell'elemento (in questo caso parti del testo) codificato.

Si potrebbe quasi affermare che lo Sgml sia un linguaggio «vuoto», riempito solo di regole per la sua stessa progettazione. Lo Sgml è dunque un» metalinguaggio» che ci fornisce le regole per realizzare un numero indefinito di linguaggi di codifica: ognuno di questi linguaggi corrisponde ad un modello di testo, o di insiemi di testo.

[6] Ricordiamo che un *Mark-up Language*, - in italiano linguaggio di marcatura o linguaggio dichiarativo – è un linguaggio nato per contrassegnare, «marcare», appunto, parti di un testo segnalando, «dichiarando», la funzione logica da esse svolta in un testo (titolo, paragrafo, verso, nota).

[7] Fra i più noti, oltre lo Sgml, ricordiamo il suo sottoinsieme Html (su cui si basa l'architettura del *World Wide Web*) e il nuovo Xml.

[8] Lanciata inizialmente nel 1987, la *Text Enconding Initiative,* TEI, è uno standard internazionale e interdisciplinare che aiuta biblioteche, musei, editori e studenti, rappresentando tutti i tipi di testi linguistici e letterari per la ricerca e l'insegnamento on line, usando uno schema di codifica che è massimamente espressivo e minimamente obsolescente. www.tei-c.org.

Se l'obiettivo finale è, dunque, creare un formato di documenti testuali leggibili dalla macchina (*Machine Readable Form*) che garantisca la massima portabilità (obiettivo finale dichiarato della Tei) deve essere garantita la condivisione del codice con cui si è codificato il testo da parte di tutti i possibili utenti. Questo scopo è raggiunto dichiarando in maniera esplicita tutti gli elementi che possono essere presenti in un documento (dai caratteri alfabetici ai marcatori utilizzati per la codifica) e le regole che governano le combinazioni di questi elementi.

Facciamo un esempio: poniamo di voler rappresentare nel nostro documento una lettera che non compare fra i primi 128 caratteri del codice Ascii [9] (che sono riconosciuti dai computer di tutto il mondo), scegliamone una largamente usata: la *e maiuscola accentata grave* (È). La sintassi che regola la costruzione delle *entity* [10] in Sgml prevede che si possa descrivere qualsiasi carattere a patto di rispettare una precisa: la *e commerciale* (&) è il segnale di inizio codifica, il *punto e virgola* (;) quello di fine codifica. Tra i due segnali è possibile inserire il carattere da codificare (nel nostro caso E), assieme a una descrizione stabilita della codifica (nel nostro caso *grave*). Tutto questo ci porta a rappresentare il segno grafico È con il gruppo: &E*grave*;. L'elenco delle *entity* legali e riconosciute sarà contenuto in un'apposita tabella che dovrà essere esplicitamente dichiarata.

[9] Codice Ascii: *American Standard Code for Information Interchange. Standard* approvato dall'Ansi (*American National Standard Institution*) alla fine degli anni Sessanta. E' uno dei codici più usati per la rappresentazione dei caratteri alfanumerici.

[10] Singola porzione di testo «marcate» da dei *tag*.

SCHEDA 7: TESTI

La maggior parte delle informazioni reperibili nella rete sono costituite da testi. Ma cosa intendiamo per «testo», quando ci riferiamo a uno scritto che appare a video navigando in rete?

In questo caso, per «testo» intendiamo un documento che fa riferimento al codice Ascii, nella sua versione a 128 o a 256 caratteri, indipendentemente dalla modalità con cui viene reso fruibile in rete.

Infatti, sono sostanzialmente due le modalità con cui si possono mettere testi su Internet: la disponibilità per la lettura e quella per il trasferimento, chiamato *download*. La prima tipologia di testi può essere letta in linea, la seconda prevede il trasferimento dalla sede remota al nostro computer e una lettura effettuata in un secondo momento, scollegati dalla rete, o più comunemente detto *off line*.

Nel secondo caso è dunque possibile utilizzare formati non compatibili con la rete, ma compatibili con un software disponibile localmente: possiamo preparare il file nel formato di un qualsiasi *word processor* e avvertire l'utente che il file è leggibile con quel determinato programma, ma possiamo anche «comprimere» il testo per diminuire le sue dimensioni e accorciare il tempo di trasmissione. In entrambi i casi il testo risulta illeggibile via Internet, ma è sempre un testo utilizzabile dopo un trattamento.

Possiamo quindi «comprimere» (in gergo *zippare*) l'intera *Divina Commedia* e metterla in un sito *Ftp*. Gli utenti interessati possono venire a sapere dell'esistenza del file nel nostro sito tramite uno dei tanti motori di ricerca disponibili su Internet, possono «scaricare» il file e leggerlo, dopo averlo «de-compresso».

La sigla *Ftp*, apparsa poco sopra, è alla base della modalità di distribuzione dei testi: sta per *File Transfert Protocol*, indica, appunto, il protocollo con cui i Internet si trasferiscono i file da un sito all'altro. Nel caso, invece, che si voglia leggere i file in linea (come accade in genere per riviste, testi elettronici e classici rivisitati) non abbiamo altra scelta che ricorrere a una codifica che ne consenta la corretta gestione da parte dei *browser* utilizzati per la navigazione.

Le codifiche disponibili sono essenzialmente due (per i quali rimandiamo alle specifiche schede), ossia *Sgml* e *Html*. Attualmente i file *Html*, più semplici da realizzare, sono più diffusi e possono essere letti per mezzo di un protocollo diverso dallo *Ftp*, parliamo dello *Http*. *Http*, che sta per *HyperTexct Transfer Protocol*, costituisce il fenomeno alla base del fenomeno Internet, in quanto rende possibile la «navigazione» della rete per mezzo dei *browser*.

SCHEDA 8: XML

L'*eXtensible Markup Language*, XML, è un linguaggio di marcatura inventato dall'organismo che stabilisce gli standard per il Web, il W3C (*World Wide Web Consortium*) [11]. I linguaggi di markup definiscono il documento, specificando come il suo contenuto deve essere interpretato.

Il più noto linguaggio di *markup*, e anche il più diffuso, è l'HTML [12] (*Hypertext Markup Language*), che si usa per creare pagine Web. Ambedue, XML e HTML, derivano dal SGML [13] (*Standard Generalized Markup Language*) un linguaggio più generico, che ha delle possibilità di applicazioni vastissime, data la sua flessibilità.

SGML è un meta-linguaggio usato per descrivere linguaggi applicativi, la sua complessità ne limita l'uso ai soli specialisti. Questo linguaggio garantisce l'accessibilità e l'interscambiabilità dei dati, per questo molte applicazioni sono oggi state tradotte da SGML in XML. XML è logica: ogni evento può essere descritto da una struttura gerarchica, che a sua volta può essere espressa in XML. Ciò che caratterizza un linguaggio di marcatura sono i *tag*, che possono essere definiti attributi personalizzati di un documento.

[11] L'indirizzo internet è <http://www.w3c.org>.
[12] Vedi Scheda n. 4.
[13] Vedi Scheda n. 6.

XML permette di creare linguaggi di markup personalizzati, o meglio indicizzare un documento categorizzandolo, in modo che dopo la sua archiviazione sia facile consultarlo. XML consente di descrivere un documento in un modo che possa essere «capito», interpretato dalla macchina. Esso fornisce un meccanismo (*Document Type Definition* [14]), che consente di condividere la struttura dei dati.

XML facilita lo scambio e la gestione dei dati, in quanto essi sono memorizzati come testo. L'evoluzione dei programmi e delle applicazioni ha complicato lo scambio dei dati, infatti spesso accade che una versione non sia in grado di leggere dati registrati in versioni più aggiornate. XML invece, in quanto linguaggio e non applicazione, non diventa desueto, e inoltre non deve essere compilato (come ad es. Pascal), ma è scritto in chiaro e può essere letto sia dalla mente umana sia da un altro linguaggio di programmazione. Ad esso si può applicare un motore di ricerca in grado di estrapolare dal testo non solo un termine, ma un intero contesto.

[14] Vedi Scheda n. 2.

Infatti in un futuro prossimo, lo studioso si troverà di fronte una enorme massa di dati codificati in XML riguardanti ogni campo dello scibile umano. A quel punto dovrà essere in grado di comprendere la struttura, il DTD, dei documenti. Inoltre esso determina un problema di tipo critico: gli indici, gli elenchi alfabetici tematici o dei nomi, che sono alla base di ogni ricerca, si trasformano da statici a dinamici.

Infatti con XML oltre ad avere immediatamente la struttura ad albero di un documento, si possono richiedere di volta in volta con un semplice motore di ricerca termini specifici. La ricerca elettronica risulta quindi essere fluida, mentre la caratteristica dell'edizione critica è la permanenza.

Un'altra tecnologia che usa linguaggio XML è l'e-book. [15]

[15] Vedi Scheda n. 3.

DOCUMENTO 1
ARTICOLO DI ROBERT DARNTON SUL PREMIO GUTENBERG-E ESTRATTO DAL SITO DELL'*AMERICAN HISTORICAL ASSOCIATION*

(http://www.theaha.org/prizes/gutenberg/rdarnton2.cfm)

What Is the Gutenberg-e Program?
By Robert Darnton

The American Historical Association established the Gutenberg-e program in 1999 in order to promote high-quality scholarly publishing on the Internet and to work toward a solution of the crisis in university presses. That crisis is poorly understood by many historians, especially older historians in well-known universities who have had no difficulty in publishing their work. Studies by the AHA and professional organizations such as the Research Library Group have demonstrated that it is almost impossible for beginning historians to get their dissertations published if they work in fields such as African history, colonial Latin America, or even early modem Europe. Moreover, the technology is evolving so rapidly that publishing of all sorts is shifting massively to the Internet. By creating a program to publish top dissertations on the Web, the AHA intends to set standards for electronic publishing in general.

Each year, the AHA conducts a national competition for the best dissertations. A panel of distinguished historians selects them, using procedures far more rigorous than those employed in most university presses. The winners receive a prize of $20,000 to be used for converting their dissertations into the best possible electronic books. Sometimes they do short stints of supplementary research and buy released time from their home universities, but they always concentrate on rewriting and on adapting their text to the new possibilities opened up by the electronic medium. They also receive extensive advice from Columbia University Press, which publishes the Gutenberg-e series and has pioneered in scholarly publishing on the Web. Columbia can draw on its own network of readers, but it has found that the reports

submitted by the panel of judges are so thorough that it has no need of further refereeing.

How will these «e-books» differ from conventional monographs? On one level, they will be very similar. Although the e-books will vary, they will contain narrative and analysis that will resemble those in the best university-press books. Their texts can be printed out and read in the normal way. Columbia University Press will present the winners with bound copies, which will look like normal books and can be submitted to tenure committees as evidence of scholarly accomplishment. But at other levels, the e-books will contain material that could never be conveyed by print: extensive documentation, hyperlinks to supplementary secondary literature, recordings, images, music, and historiographical and methodological discussions. The possibilities are endless, but the execution of the work must be of the highest quality. Columbia has the technological and editorial skill to maintain those quality standards.

If these dissertations are so good, why could their authors not publish them in the conventional manner? In some cases, that might be possible; but top presses like California and Harvard have abandoned monographs in several fields of history. Moreover, the prize-winners have the option of publishing their revised dissertations in printed form after the electronic version has been made accessible on the Web. It seems likely, however, that specialized historical scholarship will be communicated increasingly through the Internet and that new techniques of downloading, printing, and binding will open up the possibility of «instant books» and custom-made paperbacks. In a few years, readers will be able to select material that is especially relevant to their needs and to create their own books from the textual riches of the larger e-book. Gutenberg-e, as well the larger program—the History E-Book Project—sponsored by the American Council of Learned Societies, is intended to take the lead in these developments and to make sure that they conform to the highest scholarly standards. But it is not meant to replace the classic codex, which will probably continue to exist for the foreseeable future.

Finally, it should be stressed that the Gutenberg-e prizes represent the highest distinction that can be bestowed by the American Historical Association. The prizewinners, six a year, stand out as the most talented historians of their generation. The selection procedure is so rigorous that it serves as a guarantee of quality control. Following extensive coverage in the New York Times, the Chronicle of Higher Education, the New York Review of Books, and other publications, the whole program has gained recognition as a new departure in scholarship. Gutenberg-e is not a gimmick and not a technological blind alley. It is a way of promoting the best in scholarship at a time of crisis.

Robert Darnton (Princeton Univ.) is the immediate past president of the AHA.

DOCUMENTO 2
ESTRATTI DALLA TRADUZIONE ITALIANA DELLA GUIDA TEI PER LA CODIFICA DIGITALE DEI TESTI
(<HTTP://WWW.TEI-C.ORG/LITE/TEIU5_IT.HTM>)

a) Frontespizio
b) Sommario
c) Prefazione
d) Introduzione

a) Frontespizio

TEI Lite: introduzione alla codifica dei testi

Lou Burnard
C. M. Sperberg-McQueen
Documento N. TEI U5
Giugno 1995

Traduzione italiana di:
Fabio Ciotti, Guendalina Demontis, Giuseppe Gigliozzi, Massimo Guerrieri, Andrea Loreti
Revisione e cura traduzione italiana di:
Fabio Ciotti
(ciotti@axrma.uniroma1.it)
Gennaio 1998

Questo documento fornisce un'introduzione alle indicazioni elaborate dalla *Text Encoding Initiative* (TEI), descrivendo un sottoinsieme facilmente utilizzabile dell'intero schema di codifica. Lo schema qui documentato può essere utilizzato per codificare una vasta gamma di

caratteristiche testuali comunemente riscontrate, in modo da ottimizzare l'utilizzabilità delle trascrizioni elettroniche e facilitare il loro scambio fra studiosi che utilizzano diversi sistemi informatici. Esso è altresì pienamente compatibile con l'intero schema TEI, definito dal documento *TEI P3*, *Guidelines for Electronic Text Encoding and Interchange*, pubblicato a Chicago e Oxford nel maggio del 1994.

La versione originale di questo testo – in lingua inglese – può essere reperita attraverso World Wide Web agli indirizzi:

http://www.tei-c.org/Lite/teiu5_en.tei

Il documento è anche disponibile in formato HTML sui siti:

http://www.tei-c.org/Lite/

La *Document Type Definition* SGML formale qui descritta, si trova negli stessi siti, nel file *teilite. dtd*:

http://www.tei-c.org/Lite/DTD/teilite.dtd

b) Sommario

- 10 Interventi editoriali
- 11 Omissioni, soppressioni e aggiunte
- 12 Nomi, date, numeri e abbreviazioni
    - 12.1 Nomi ed espressioni referenziali
    - 12.2 Date e orari
    - 12.3 Numeri
    - 12.4 Abbreviazioni e loro espansioni
    - 12.5 Indirizzi
- 13 Liste
- 14 Citazioni bibliografiche
- 15 Tavole e Tabelle
- 16 Immagini e grafica
- 17 Interpretazione ed analisi
    - 17.1 Frasi ortografiche
    - 17.2 Elementi generici di interpretazione
- 18 Documentazione tecnica
    - 18.1 Elementi supplementari per i documenti tecnici.
    - 18.2 Sezioni generate automaticamente
    - 18.3 Generazione di indici
- 19 Set di caratteri, diacritici, etc.
- 20 Elementi dell'avantesto ed annessi
    - 20.1 Elementi dell'avantesto
        - 20.1.1 Frontespizio
        - 20.1.2 Materiali introduttivi
    - 20.2 Elementi degli annessi
        - 20.2.1 Divisioni strutturali degli annessi
- 21 Il frontespizio elettronico
    - 21.1 Descrizione del file
        - 21.1.1 Dichiarazione del titolo
        - 21.1.2 Dichiarazione dell'edizione
        - 21.1.3 Dichiarazione della dimensione
        - 21.1.4 Dichiarazione della pubblicazione
        - 21.1.5 Dichiarazione di collane e note
        - 21.1.6 Descrizione della fonte

c) Prefazione

PREFAZIONE ALLA EDIZIONE ITALIANA

Questo documento è la traduzione italiana del documento ufficiale della Text Encoding Initiative (TEI) numero TEI U5 (Giugno 1995), TEI Lite: An Introduction to Text Encoding for Interchange, redatto da Lou Burnard e Michael Sperberg-McQueen.

Esso fornisce la documentazione di uno schema di codifica SGML (e della relativa Document Type Definition), comunemente noto come TEI Lite, che a sua volta costituisce una versione semplificata dell'intero schema di codifica definito dalla TEI e documentato nel testo TEI P3, Guidelines for Electronic Text Encoding and Interchange (che in questa sede chiameremo Norme TEI o, più semplicemente, TEI P3)

Lo schema di codifica della TEI, basato sulla sintassi dello Standard Generalized Markup Language (SGML, ISO 8879), è indirizzato a tutti coloro che intendono produrre e diffondere testi in formato elettronico a fini scientifici e di ricerca, in particolare nel dominio umanistico. Esso consente infatti di rappresentare la struttura astratta di varie tipologie testuali (testo in prosa, testo poetico, testo drammaturgico,

fonte manoscritta, etc.), e le caratteristiche testuali rilevanti per diverse aree di ricerca (filologia, analisi linguistica, tematica, narratologica, etc.).

Il sottoinsieme denominato TEI Lite, qui documentato, è stato sviluppato al fine di facilitare l'applicazione dello schema da parte degli utenti senza richiedere lo studio dell'intera DTD, anche nelle sua parti più esoteriche. Esso permette la creazione di documenti TEI-compliant (compatibili, cioè, con l'intero schema) in maniera rapida, e si presta facilmente allo sviluppo di applicazioni.

Attualmente le maggiori istituzioni di ricerca a livello mondiale nel campo informatico umanistico utilizzano la TEI per la creazione di banche dati testuali di ricerca. La complessità, l'estensibilità e la diffusione, unitamente alla sua origine ed evoluzione interna all'ambito umanistico, ne fanno infatti il più valido strumento di codifica per la creazione di testi elettronici, sia a puro fine editoriale, che come supporto per l'analisi informatizzata dei testi. Alcuni testi della Letteratura Italiana codificati in formato SGML/TEI sono disponibili presso il sito Web del Centro Ricerche Informatica e Letteratura (CRILet), all'indirizzo http://crilet.let.uniroma.it, dove è disponibile anche una versione HTML del presente documento.

Questa traduzione è il frutto del lavoro (del tutto volontario) di Fabio Ciotti, Guendalina Demontis, Giuseppe Gigliozzi, Massimo Guerrieri, Andrea Loreti. La versione finale è stata revisionata e curata da Fabio Ciotti, che si assume tutte le responsabilità per la traduzione dei termini tecnici e speciali.

L'originale è stato esaminato nella versione SGML/TEI distribuita presso il sito ufficiale della Text Encoding Initiative nel file «teiu5. tei» (http://www.tei.uic.edu/orgs/tei/intros/teiu5.tei).

Nella traduzione si è cercato di rimanere quanto più possibile aderenti al testo inglese. Come detto Guidelines è stato tradotto con Norme. Per quanto riguarda gli esempi, laddove è stato ritenuto desiderabile (e, soprattutto possibile), si sono inseriti testi originali della tradizione letteraria italiana. Per quanto attiene ai termini tecnici ed alla terminologia SGML, si è preferito in linea di massima tradurli in lingua italiana in base al seguente schema:

generic identifier = identificatore generico
tag = marcatore

element = elemento
attribute = attributo
unique identifier = identificatore unico
entity = entità
entity name = nome di entità

Anche i termini tecnici specifici della TEI sono stati generalmente tradotti. In particolare, in riferimento allo statuto del valore degli attributi legal è stato tradotto con «permessi», suggested o possible con «consigliati» «possibili», «esemplificativi» o «suggeriti»; global attributes è stato tradotto come «attributi globali». I valori suggeriti per gli attributi sono stati generalmente tradotti o indicati in italiano (con alcune eccezioni).

Il presente documento è disponibile in formato HTML ai seguenti indirizzi:
http://crilet.let.uniroma1.it/sgml/teiu5-it/teu5-it.html
http://rmcisadu.let.uniroma1.it/crilet/sgml/teiu5-it/teiu5-it.html

Una versione suddivisa in più file per facilitarne la consultazione online è disponibile ai seguenti indirizzi:
http://crilet.let.uniroma1.it/sgml/teiu5-it/split/teu5-it.html
http://rmcisadu.let.uniroma1.it/crilet/sgml/teiu5-it/split/teiu5-it.html

d) Introduzione

Le Norme (Guidelines) della Text Encoding Initiative (TEI) sono indirizzate a tutti coloro che intendono scambiare informazioni archiviate in formato elettronico. Esse sottolineano l'importanza dello scambio di informazioni testuali, ma trattano anche di altre forme di informazioni (quali immagini e suoni). Le Norme sono applicabili indifferentemente sia per la creazione di nuove risorse che per lo scambio di quelle già esistenti.

Le Norme forniscono un mezzo per rendere esplicite certe caratteristiche di un testo in modo tale da facilitarne il trattamento mediante programmi di computer basati su diverse piattaforme. Definiamo questo processo di esplicitazione marcatura (markup) o codifica (encoding). Qualsiasi rappresentazione di un testo su un computer usa una

qualche forma di codifica; la TEI è stata creata sia per ovviare alla eccessiva varietà di schemi di codifica tra loro incompatibili che ostacolano la ricerca scientifica, sia per il crescente numero di applicazioni scientifiche che ora vengono individuate per i testi in formato elettronico.

Le Norme TEI utilizzano lo Standard Generalized Markup Language (SGML) per definire il proprio schema di codifica. SGML è uno standard internazionale (ISO 8879), sempre più usato nell'industria dell'informazione, che permette la definizione formale di uno schema di codifica in termini di elementi e attributi, e di regole che gestiscano la loro occorrenza all'interno di un testo. L'applicazione dello SGML da parte della TEI è ambizioso nella sua complessità e generalità, ma fondamentalmente non differisce da quella di qualsiasi altro schema di codifica SGML; conseguentemente qualsiasi applicazione SGML generica è in grado di elaborare testi conformi alla TEI.

La TEI è sponsorizzata dall'Association for Computers and the Humanities, dall'Association for Computational Linguistics, e dall'Association for Literary and Linguistic Computing. Finanziamenti sono stati in parte forniti dall'U. S. National Endowment for the Humanities, Directorate General XIII of the Commission of the European Communities, dall'Andrew W. Mellon Foundation, e dal Social Science and Humanities Research Council of Canada. Le Guidelines sono state pubblicate nel maggio 1994, dopo sei anni di sviluppo che ha coinvolto parecchie centinaia di studiosi di tutto il mondo provenienti da diversi ambiti accademici.

Gli obiettivi principali della TEI sono stati definiti, agli inizi del suo lavoro, nelle dichiarazioni conclusive di una conferenza programmatica tenutasi al Vassar College, New York, nel novembre 1987; questi Poughkeepsie Principles sono stati ulteriormente elaborati in una serie di documenti progettuali. Le Norme, come affermano tali documenti, dovrebbero:

- essere in grado di rappresentare le caratteristiche testuali necessarie per la ricerca;
- essere semplici, chiare e concrete;
- essere di semplice utilizzazione per i ricercatori senza il ricorso a software specializzati;

- permettere una definizione rigorosa e un'efficiente elaborazione dei testi;
- consentire estensioni definite dall'utente;
- Il mondo della ricerca è vasto e variegato. Affinché le Norme godessero della più ampia accoglienza, è stato importante assicurare che:
    - l nucleo comune delle caratteristiche testuali fosse facilmente condiviso;
    - le caratteristiche specialistiche supplementari fossero facili da aggiungere (o da rimuovere) da un testo;
    - fossero possibili molteplici codifiche parallele della stessa caratteristica;
    - la ricchezza della codifica potesse essere definita dall'utente, con una soglia minima molto bassa;
    - fosse fornita un'adeguata documentazione del testo e della sua codifica.

Il presente documento descrive una selezione, di facile utilizzazione, dell'esteso insieme di elementi SGML e suggerimenti definiti dalla TEI in conformità con tali obiettivi progettuali, denominata TEI Lite.

Selezionando tra le diverse centinaia di elementi SGML definiti dallo schema completo della TEI, abbiamo cercato di identificare un utile «insieme di partenza», comprendente gli elementi che quasi ogni utente dovrebbe conoscere. L'esperienza fatta lavorando con la TEI Lite sarà di inestimabile valore per comprendere l'intera DTD TEI, e per sapere quali parti opzionali della DTD siano necessarie per lavorare su particolari tipi di testo.

I nostri obiettivi, nel definire questo sottoinsieme, possono essere riassunti nel modo seguente:

- esso dovrebbe includere la maggior parte dell'insieme «fondamentale» di marcatori della TEI, dal momento che questo contiene elementi rilevanti virtualmente per tutti i testi e per tutti i tipi di elaborazione testuale;
- esso dovrebbe essere in grado di trattare adeguatamente una varietà di testi ragionevolmente ampia, al livello di dettaglio incontrato nella pratica già esistente;

- esso dovrebbe essere utile sia per l'elaborazione di nuovi documenti sia per la codifica di quelli già esistenti;
- esso dovrebbe essere utilizzabile con un ampio spettro di applicazioni SGML già esistenti;
- esso dovrebbe essere derivabile dall'intera DTD TEI, usando il meccanismo di estensione descritto nelle Norme;
- esso dovrebbe essere tanto conciso e semplice quanto consentito dalla conformità agli altri obiettivi.

Il lettore potrà da solo giudicare il nostro successo nel far fronte a questi obiettivi. Nel momento in cui scriviamo, la nostra fiducia di aver almeno parzialmente raggiunto i nostri scopi nasce dal suo uso nella pratica di codifica di testi reali. L'Oxford Text Archive usa la TEI Lite per tradurre i testi del suo patrimonio dai loro schemi originali di codifica in formato SGML; gli Electronic Text Centers della University of Virginia e della University of Michigan, hanno usato la TEI Lite per codificare il loro patrimonio. E la Text Encoding Initiative stessa, utilizza la TEI Lite nella sua documentazione tecnica corrente – incluso questo documento.

Sebbene abbiamo cercato di rendere questo documento autosufficiente, come si addice a un testo didattico e introduttivo, il lettore dovrebbe essere consapevole che esso non copre ogni dettaglio dello schema di codifica TEI. Tutti gli elementi qui descritti sono documentati compiutamente nel testo completo delle Norme, che dovrebbe essere consultato per avere informazioni di riferimento autorevoli su questi, e sui molti altri che non sono qui descritti. Si presuppone, inoltre, una conoscenza di base dello SGML.

DOCUMENTO 3
RELAZIONE TECNICA DEL «FILARETE ON LINE» ESTRATTA DAL SITO UFFICIALE
(<HTTP://WWW.UNIMI.IT/ATENEO/FILARETE/RELAZIONE.HTM>)

La direzione della Collana «Il Filarete» – Pubblicazioni della Facoltà di Lettere e Filosofia ha deciso di rendere nuovamente disponibili alcuni dei volumi della serie, da tempo esauriti e tuttavia ancora richiesti da studiosi italiani e stranieri.

Scartata l'ipotesi di ristamparli, si è pensato di digitalizzarli e di pubblicarli nella sezione del sito dell'Università dedicato appunto alla Collana; parte integrante del progetto è anche l'intenzione di mantenersi fedeli, per quanto riguarda l'aspetto editoriale, allo spirito della nuova serie (che ha preso il via nel 2000 con la collaborazione della casa editrice LED). La nuova serie è stata progettata

con l'intento non solo di continuare a diffondere, insieme alla rivista «Acme», i migliori risultati della ricerca scientifica che si svolge nella Facoltà di Lettere e Filosofia, ma anche, in un'epoca in cui l'editoria elettronica facilita la produzione di libri di modesta, quando non pessima, qualità editoriale, con il proposito di offrire un piccolo ma significativo esempio di come la più moderna tecnologia possa essere utilmente messa al servizio della migliore tradizione dell'arte della stampa.

Diverse iniziative, sia nazionali sia internazionali, hanno come scopo la digitalizzazione di volumi in precedenza pubblicati a stampa. Fra quelle più importanti c'è sicuramente Gallica, che permette l'accesso gratuito a 70.000 opere digitalizzate, 80.000 immagini e molte ore di risorse sonore. L'intento è quello di offrire una sorta di biblioteca patrimoniale ed enciclopedica a carattere multidisciplinare, prevalentemente francofona, ma che comprende anche classici stranieri in versione originale o in traduzione. Come si legge nella presentazione: «questo insieme di romanzi, saggi, periodici, testi celebri e opere più rare è qui riunito per permettere ad ogni tipo di lettore, dal curioso al bibliofilo, dal liceale all'universitario, di approfondire la conoscenza di un'epoca nei suoi aspetti politici, filosofici, scientifici o letterari».

Le opere che Gallica mette a disposizione sono prevalentemente

in formato immagine, sono in altre parole delle fotografie delle diverse pagine di cui è composto un volume, fotografie che è possibile sfogliare attraverso l'uso di programmi appositi. Questo, se da un lato assicura un'assoluta fedeltà all'originale, è una rinuncia a sfruttare le possibilità offerte dalle edizioni digitali.

Una scelta analoga nella finalità generale, ma opposta per quanto riguarda l'aspetto tecnico, è stata compiuta da Liber Liber. L'Associazione, attraverso il Progetto Manuzio,

ha l'ambizione di concretizzare un nobile ideale: la cultura a disposizione di tutti. Come? Capolavori della letteratura, manuali, tesi di laurea, riviste e altri documenti in formato elettronico disponibili sempre, in tutto il mondo, a costo zero e con accorgimenti tecnici tali da garantirne la fruibilità anche a non vedenti e altri portatori di handicap.

I formati utilizzati sono HTML (HyperText Markup Language) per la versione ipertestuale del libro, che può includere immagini o suoni; RTF (Rich Text Format), un formato compatibile con la maggior parte dei programmi di videoscrittura e quindi particolarmente adatto nel caso si voglia stampare il testo; solo testo (non ci sono immagini, gli stili come il corsivo, il neretto, ecc. non sono presenti).

Nel caso di Gallica, quindi, una riproduzione dell'originale in formato TIFF (Tag-based Image File Format) o PDF (Portable Document Format), nel caso di Liber Liber, invece, la disponibilità del testo, ma presentato in maniera diversa rispetto all'originale.

Per quanto riguarda le opere da includere nella serie on line del «Filarete», trattandosi di studi critici e di edizioni di testi, si è deciso di percorrere una strada che consentisse di recuperare il testo e di generare, a partire da questo, sia una versione che ne renda possibile la consultazione in linea, sia una versione che ne riproduca esattamente l'aspetto editoriale: questa scelta ha l'immediato vantaggio di consentire agli studiosi di citare l'opera dall'edizione elettronica senza dover consultare quella cartacea, ma anche quello, non trascurabile, di conservare la memoria delle tipologie editoriali adottate.

Posto che, per poter generare un file HTML e un file PDF a partire dal testo, occorre che questo sia marcato, si trattava di valutare quale di linguaggio di marcatura, fra quelli disponibili, poteva rispondere meglio alle esigenze e alle finalità del progetto.

Tali finalità ci sono sembrate corrispondere pienamente con quanto è stato recentemente sostenuto nel White paper *XML and PDF: Why We Need Both* di Impressions, un'azienda specializzata in editoria elettronica, e cioè che

> It's becoming clearer every day that the ideal workflow for book and journal publishers starts with XML or SGML, which can then be used to generate HTML (or other forms of XML) for Web publishing and PDF for electronic versions of pages to be printed (whether they're printed in bulk or on demand).

I cenni che seguono e gli esempi offerti saranno sufficienti ad illustrare sinteticamente sia alcuni degli aspetti tecnici sia la natura delle scelte operate e gli orientamenti per quelle future.

Come è noto, qualsiasi documento creato a partire da un linguaggio di marcatura è costituito da contenuto e da marcatori. Scopo dei marcatori è quello di veicolare informazioni (strutturali, relative alla presentazione, semantiche) che vanno oltre il contenuto. La caratteristica peculiare dei linguaggi di marcatura sta appunto nel fatto che le informazioni aggiuntive sono inserite all'interno dei documenti stessi.

Ecco un frammento di un documento HTML – il più diffuso tra questi linguaggi – tratto dalla *HTML 4.01 Specification* e composto da un titolo di secondo livello (delimitato dai marcatori `<h2>` e `</h2>`) e da un paragrafo (`<p>` e `</p>`):

```
<h2>Abstract</h2>
<p>This specification defines the HyperText
Markup Language (HTML), the publishing language
of the World Wide Web. </p>
```

La descrizione di un documento che HTML consente di fare non può tuttavia soddisfare chi voglia descrivere esattamente la struttura di un documento. Manca, tanto per fare un esempio particolarmente significativo, un elemento che permetta di contrassegnare la fine di una pagina.

Invece di rivolgerci ad un linguaggio composto da un numero finito di elementi, conviene piuttosto utilizzare un metalinguaggio, ov-

vero un linguaggio che può descrivere altri linguaggi e che quindi ci consentirà di definire esattamente gli elementi di cui abbiamo bisogno per strutturare i nostri testi.

XML (*eXtended Markup Language*) è questo metalinguaggio. Esso non offre un insieme di elementi predefinito, come HTML, ma permette di crearne a piacimento. Inoltre, attraverso la Document Type Definition (DTD), consente la convalida dei documenti: è possibile cioè rendere obbligatoria la presenza di alcuni elementi all'interno del documento rendendone altri facoltativi e di imporre delle restrizioni riguardo l'ordine in cui i diversi elementi devono comparire.

Nel nostro caso, la scelta si è orientata verso quell'insieme di elementi definito dalla *TEI P4. Guidelines for Electronic Text Encoding and Interchange*. Nata nel 1987, la Text Encoding Initiative (TEI), è uno standard internazionale e interdisciplinare capace di consentire a biblioteche, musei, editori e singoli studiosi di riprodurre ogni genere di testi letterari e linguistici per la ricerca in linea e per l'insegnamento servendosi di uno schema di codifica che unisca il massimo di espressività al minimo di obsolescenza. A partire dal dicembre 2000, si è costituito un consorzio internazionale per sostenere e sviluppare la TEI.

Per comprenderne la funzionalità, basterà in questa sede un esempio, tratto dalla traduzione italiana della *TEI Lite: An Introduction to Text Encoding for Interchange* di Lou Burnard e C. M. Sperberg-McQueen.

Un ipotetico operatore trascrive, servendosi di un programma di videoscrittura**,** un testo stampato (nel nostro caso si tratta di un brano del *Lanciatore di giavellotto* di Paolo Volponi) restando, per quanto è possibile, fedele all'originale: conserva le interruzioni di riga originali, introduce spazi per rappresentare l'impostazione tipografica dei titoli e dei salti di pagina, e così via.

Capitolo 16 163
– Sono contenta che tu sia bravo, – infine poté dire quietamente la madre.
Damìn sorrise. – Bravo tanto da impressionare i professori. Tanto bravo da diventare un artista. Te lo meriti perché sei buono; e se lo merita anche tuo nonno. Anche lui è un artista; anche se è rimasto a fare cocci… un vero artista. Chissà anche lui come sarà contento. Il segno D. P. continua; continua anche nell'arte, come ha detto il professore.
– Ma io andrò via da Fossombrone, – disse Damìn, – appena potrò; appena sarò più grande. Non ne posso più: per fortuna questa scuola… se no sarei già scappato da tempo. Mi sento soffocare a Fossombrone. E ancora di più dentro casa.
– E perché vorresti fuggire? Cos'è che ti fa soffocare? –- la madre esitava nella domanda, ma la sincerità della pena la costringeva a parlare. – Potrai sì andar via, quando sarai più grande, per il tuo lavoro. E allora sarà giusto, anche se difficile da sopportare. Viene sempre purtroppo il momento in cui i figli debbono lasciare i genitori; specie nei paesi, e specie se i figli sono bravi e hanno studiato.
– E tu ci hai mai lasciato? – domandò Damìn.
– Io? Lasciato? E perché? Quando avrei potuto lasciarvi io?
Damìn stringeva il tavolino e si abbassò per andare a guardarne le gambe.
164
– Vi ho dato forse l'impressione che avrei potuto lasciarvi? Io? Lasciarvi, tu e Lavinia? Dove avrei potuto mai andare?

Salta subito all'occhio che questa trascrizione presenta una serie di problemi:

- i numeri delle pagine e i titoli passanti sono inframmezzati al testo;
- non c'è distinzione tra la virgoletta singola e gli apostrofi, cosicché è difficile distinguere le parole tronche da quelle apostrofate;
- non c'è distinzione tra i segni che introducono un discorso diretto e la linea di sillabazione;
- le divisioni del paragrafo sono segnate con uno spazio bianco e so-

> no stati introdotti ritorni a capo alla fine di ogni riga. Se dovesse cambiare anche solo la dimensione del carattere usato per stampare il testo, la formattazione diventerebbe inevitabilmente problematica.

Ecco lo stesso brano codificato secondo le regole stabilite dalla TEI:

```
<div1 type=capitolo n='16'>
<p>&mdash; <q>Sono contenta che tu sia bravo, </q> &mdash; infine
pot&eacute; dire quietamente la madre.
<p>Dam&iacute; n sorrise.
<p>&mdash; <q>Bravo tanto da impressionare i professori. Tanto
bravo da diventare un artista. Te lo meriti perch&eacute; sei
buono; e se lo merita anche tuo nonno. Anche lui &egrave; un
artista; anche se &egrave; rimasto a fare cocci… un vero
artista. Chiss&agrave; anche lui come sar&agrave; contento. Il
segno D. P. continua; continua anche nell'arte, come ha detto il
professore. </q>
<p>&mdash; <q>Ma io andr&ograve; via da Fossombrone, </q>
&mdash;
disse Dam&iacute; n, &mdash; <q>appena potr&ograve;; appena
sar&ograve; pi&uacute; grande. Non ne posso pi&uacute;: per fortuna
questa scuola… se no sarei gi&agrave; scappato da tempo.
Mi sento soffocare a Fossombrone. E ancora di pi&uacute; dentro
casa. </q>
<p>&mdash; <q>E perch&eacute; vorresti fuggire? Cos'&egrave; che ti
fa soffocare?</q> &mdash; la madre esitava nella domanda, ma la
sincerit&agrave; della pena la costringeva a parlare. &mdash;
<q>Potrai s&iacute; andar via, quando sarai pi&uacute; grande, per
il tuo lavoro. E allora sar&agrave; giusto, anche se difficile da
sopportare. Viene sempre purtroppo il momento in cui i figli
debbono lasciare i genitori; specie nei paesi, e specie se i figli
sono bravi e hanno studiato. </q>
<p>&mdash; <q>E tu ci hai mai lasciato?</q> &mdash; do-
mand&ograve;
Dam&iacute; n.
<p>&mdash; <q>Io? Lasciato? E perch&eacute; ? Quando avrei potuto
lasciarvi io?</q>
<p>Dam&iacute; n stringeva il tavolino e si abbass&ograve; per
andare a guardarne le gambe.
<pb n='164'>
```

&mdash <q>Vi ho dato forse l'impressione che avrei potuto lasciarvi? Io? Lasciarvi, tu e Lavinia? Dove avrei potuto mai andare?</q>

In questo modo:

- le divisioni tra paragrafi sono contrassegnate esplicitamente;
- gli apostrofi sono distinti dalle virgolette singole;
- riferimenti ad entità sono usate per le lettere accentate e per i trattini lunghi;
- le divisioni delle pagine sono state marcate con l'elemento <pb>;.

Per semplificare la ricerca e il trattamento automatico, la divisione in righe dell'originale non è stata mantenuta e le parole spezzate per motivi tipografici alla fine di una linea sono state riunite. Ma se si fosse voluto mantenere la divisione in righe dell'originale, questa avrebbe potuto essere registrata.

Passando ora alla messa in atto del progetto, le fasi di lavorazione da prevedere allorché tutta la procedura sarà definitivamente messa a punto sono le seguenti:

- il riconoscimento ottico dei caratteri;
- la correzione;
- la marcatura del testo, in modo da conservarne la struttura;
- la generazione, a partire dal testo così marcato, di un file PDF che riproduca esattamente l'originale (utilizzando lo stesso formato pagina e la stessa famiglia di caratteri) e di un file HTML per la consultazione in linea.

Poiché un progetto di questa complessità richiede che le diverse fasi vengano messe a punto con estrema cura, è parso opportuno valutare per prima cosa i tempi per il riconoscimento ottico dei caratteri e per la successiva correzione. Una volta disponibile il testo, è stata utilizzata una applicazione di Desktop Publishing (Quark XPress) per impaginarlo esattamente come l'originale e generare un file PDF.

Il risultato di questa fase iniziale è rappresentato da i volumi di Dino Formaggio (*Fenomenologia della tecnica artistica*, Milano, Nuvoletti, 1953), con nuova prefazione di Gabriele Scaramuzza, e di Sergio Antonielli (*Giuseppe Parini*, Firenze, La Nuova Italia, 1973), con nuova presentazione di Alberto Cadioli ed Edoardo Esposito.

Si è poi deciso di passare alla seconda, più complessa fase del lavoro (marcatura XML, generazione dei file PDF e HTML).

Mentre marcare il testo secondo le specifiche è semplice, in quanto esse sono molto dettagliate ed esistono delle guide che contengono le risposte a tutte le domande, più complesso risulta passare dal file XML al file PDF e al file HTML. Una strada possibile – non l'unica, come vedremo subito – è quella di utilizzare un programma come Adobe FrameMaker 7.0 per costruire un'applicazione che, utilizzando la DTD TEI, permetta formattare il testo in maniera identica all'originale.

La cosa era già stata fatta, utilizzando una versione precedente del programma e per un'altra tipologia di documenti, da Pim van der Eijk. Il suo articolo, *Editing TEI documents using FrameMaker+SGML*, contiene una serie di indicazioni preziose.

Questo procedimento è stato utilizzato per il volume di G. Preti, *Fenomenologia del valore* (Messina-Milano, Principato, 1942), con prefazione di Renato Pettoello.

In futuro, il passaggio attraverso FrameMaker potrebbe essere evitato mettendo a punto una serie di fogli stile che consentano di generare un documento XSL-FO (*eXtensible Stylesheet Language-Formatting Objects*). Da questo sarebbe infatti possibile ottenere documenti PDF.

Un altro sviluppo in corso di studio prevede la generazione di documenti compatibili leggibili attraverso dai browser vocali secondo quanto stabilito dal gruppo di lavoro Voice Browser.

RIFERIMENTI BIBLIOGRAFICI

Lou Burnard, C.M. Sperberg-McQueen, *TEI Lite: An Introduction to Text Encoding for Interchange*, <http://www.tei-c.org/Lite/> (la traduzione italiana della prima edizione è disponibile all'indirizzo <http://www.tei-c.org/Lite/teiu5_it.htm>).

Danilo Deana, *Linguaggi di marcatura e fogli stile*, Milano, CUEM, 2002.

Pim van der Eijk, *Editing TEI documents using FrameMaker+SGML*, <www. sonnenglanz. net/fmtei.html>.

Extensible Markup Language (XML) 1.0 (Second Edition), edited by Tim Bray, Jean Paoli, C.M. Sperberg-McQueen, Eve Maler, <www. w3.org/TR/2000/REC-xml-20001006>.

The TEI Consortium, *TEI P4. Guidelines for Electronic Text Encoding and Interchange. XML-compatible edition,* edited by C M Sperberg-McQueen and Lou Burnard, XML conversion by Syd Bauman, Lou Burnard, Steven DeRose, and Sebastian Rahtz, <http://www.tei-c.org/Lite/teiu5_it.htm>.

XML and PDF: Why We Need Both. An Introduction to the Two Keys Technologies for Electronic Publishing,
<http://www.impressions.com/resources_pgs/SGML_pgs/XML_PDF.html>

BIBLIOGRAFIA
E RISORSE DIGITALI

BIBLIOGRAFIA

AA. VV., *Il laboratorio di storia. Problemi e strategie per l'insegnamento nella prospettiva dei nuovi curricoli e dell'autonomia didattica*, Milano, Edizioni Unicopoli, 2001.

T. Ballarino, *Internet nel mondo della legge: in appendice le norme Uncitral sul commercio e sulla firma elettronica, la legge tedesca su Internet, la storia di Internet, glossario di Internet*, Padova, CEDAM, 1998.

T. Berners-Lee, *L'architettura del nuovo web*, Milano, Feltrinelli, 2001.

G. Bettetini – B. Gasparini – N. Vittadini, *Gli spazi dell'ipertesto*, Milano, Bompiani, 1999.

G. Blasi, *Internet: storia e futuro di un nuovo medium*, Milano, Guerini studio, 1999.

S. Bordini, *Uno strabismo storiografico: il medievista e l'internet*, in R. Greci (a cura di), *Medioevo in rete tra ricerca e didattica*, Bologna, Edizioni CLUEB, 2002.

A. Cadioli, *Dall'editoria moderna all'editoria multimediale*, Milano, Edizioni Unicopoli, 1999.

M. Calvo – F. Ciotti, G. Roncaglia, *Internet 2000. Manuale per l'uso della rete*, Roma-Bari, Laterza, 1999.

F. Ciotti, *La comunicazione telematica e le discipline umanistiche*, in R. Greci (a cura di), *Medioevo in rete tra ricerca e didattica*, Bologna, Edizioni CLUEB, 2002.

F. Ciotti – G. Roncaglia, *Il mondo digitale. Introduzione ai nuovi media*, Bari, Editori Laterza, 2000.

J. Conklin, *Hypertext: An introduction and survey*, in «Computer Magazine», 1987, tr.it. in «Informatica Oggi», 1988.

P. Ferri, *La rivoluzione digitale e le nuove modalità di trasmissione del sapere*, in G. Montecchi (a cura di), *La città dell'editoria. Dal libro tipografico all'opera digitale 1880-2020*, Milano, Skira, 2001 (Catalogo della Mostra tenuta a Milano nel 2001).

G. Gigliozzi, *Il testo e il computer. Manuale di informatica per gli studi letterari*, Milano, Bruno Mondadori, 1997.

A. Gisolfi, *Ipertesto: libertà e controllo della navigazione*, in F. Bocchi, P. Denley (a cura di), *Storia e Multimedia. Atti del Settimo Congresso Internazionale dell'Association for Historical&Computing*, Bologna, Grafis Edizioni, 1994.

R. Greci, Introduzione a R. Greci (a cura di), *Medioevo in rete tra ricerca e didattica*, Bologna, Edizioni CLUEB, 2002.

K. Hafner - Matthew Lyon, *Where wizards stay up late*, trad.it. in *La storia del futuro: le origini di Internet*, Milano, Feltrinelli, 1998.

G. P. Landow, *Hypertext 2.0. The Convergence of Contemporary Critical Theory and Technology*, trad. it. in P. Ferri (a cura di), *L'ipertesto. Tecnologie digitali e critica letteraria*, Milano, Edizioni Bruno Mondadori, 1998.

P. Lévy, *Le tecnologie dell'intelligenza*, Bologna, Synergon, 1992.

G. P. Lotito, *La «leggerezza» del mondo digitale. Il sapere viaggia con i bit*, in G. Montecchi (a cura di), *La città dell'editoria. Dal libro tipografico all'opera digitale 1880-2020*, Milano, Skira, 2001 (Catalogo della Mostra tenuta a Milano nel 2001).

B. Longo, *La Nuova Editoria*, Milano, Editrice Bibliografica, 2001.

G. Mauri, *La struttura degli ipertesti*, in Mario Ricciardi (a cura di), *Oltre il testo: gli ipertesti*, Milano, Franco Angeli, 1994.

H.M. McLuhan, *Understanding Media*, New York, McGraw-Hill Book Company, 1964, trad. it. in Ettore Capriolo (a cura di), *Gli strumenti del comunicare*, Milano, il Saggiatore, 1967.

T.H. Nelson, *Come penseremo*, in J. Nyce, P. Kahn, *Da Memex a Hypertext*, Padova, Franco Muzzio Editore, 1992.

P. Ortoleva, *Mass media: dalla radio alla rete*, Firenze, Giunti, 2001.

P. Ortoleva, *Presi nella rete? Circolazione del sapere storico e tecnologie informatiche*, in S. Soldani – L. Tomassini (a cura di) *Lo storico e il computer*, Milano, Edizioni Scolastiche Bruno Mondadori, 1996.

M. Schirru, *L'ipertesto applicato alla ricerca storica. Il caso della relazione (1737-1738) del marchese di Rivarolo, viceré di Sardegna*, in F. Bocchi, P. Denley (a cura di), *Storia e Multimedia. Atti del Settimo Congresso Internazionale dell'Association for Historical&Computing*, Bologna, Grafis Edizioni, 1994.

S. Soldani – L. Tomassini, Introduzione a S. Soldani – Tomassini (a cura di), *Lo storico e il computer*, Milano, Edizioni Scolastiche Bruno Mondadori, 1996.

M. Wolf, *Gli effetti sociali dei media*, Milano, Bompiani, 1992.

A. Zorzi, *Comunicazione del sapere ed editoria digitale: problemi e prospettive per gli studi medievali*, in R. Greci (a cura di), *Medioevo in rete tra ricerca e didattica*, Bologna, Edizioni CLUEB, 2002.

RISORSE DIGITALI

F. Anania, *Internet, la storia, il pubblico*, in Serge Noiret (a cura di), *Internet e il mestiere di Storico*. Il testo dell'intervento di Anania all'indirizzo: <http://www.sissco.it/dossiers/internet/anania-sem-apr-2000.html>

M. Ansani, *La tradizione disciplinare fra innovazione e nemesi digitale*, in *Medium-evo. Gli studi medievali e il mutamento digitale*, I Workshop nazionale di studi medievali e cultura digitale, Firenze, 21-22 giugno 2001. Il testo integrale all'indirizzo: <http://www.storia.unifi.it/_PIM/ medium-evo/>

M. Ansani, *Una leggerezza complicata*, in G. Abbatista e A. Zorzi (a cura di), *Il documento immateriale. Ricerca storica e nuovi linguaggi*. Testo disponibile all'indirizzo: <http://lastoria.unipv.it/dossier/ansani.htm >

E. Baldassarri, *La comunicazione storica nell'era multimediale*, in *IS-Internet e Storia. 1° Forum telematico 15 gennaio-15 marzo 2003*. Il testo integrale di questo intervento è scaricabile, in formato pdf, all'indirizzo: <http://www.storiadelmondo.com/6/baldassarri.comunicazione.pdf>

P. Corrao, *Saggio storico, forma digitale: trasformazione o integrazione?*, in *Medium-evo. Gli studi medievali e il mutamento digitale*, I Workshop nazionale di studi medievali e cultura digitale, Firenze, 21-22 giugno 2001. Il testo integrale all'indirizzo: <http://www.storia.unifi.it/_PIM/medium-evo/>

P. Corrao, *Un dominio signorile nella Sicilia tardomedievale*, saggio ipertestuale che si può esplorare all'indirizzo: <http://www.rm.unina.it/Rivista1/venti>

R. Darnton, *An early information Age*, saggio ipertestuale disponibile all'indirizzo: <http://www.indiana.edu/~ahr/darnton>

R. Darnton, *The New Age of the Book*, «New York Review of Books», Volume 46, Number 5, March 18, 1999. Disponibile all'inidirizzo: <http://www.nybooks.com/nyrev/WWWarchdisplay.cgi?19990318005F>

R. Delle Donne, *Gli strumenti di consultazione*, in *Medium-evo. Gli studi medievali e il mutamento digitale*, I Workshop nazionale di studi medievali e cultura digitale, Firenze, 21-22 giugno 2001. Il testo integrale all'indirizzo:

<http://www.storia.unifi.it/_PIM/medium-evo

M.A. Garcia, *Testo e ipertesto*, Roma, 23 gennaio 1996, intervista a *Mediamente*. Il testo integrale all'indirizzo:
<http://www.mediamente.rai.it/home/bibliote/biografi/g/garcia.htm>

David Kolb, *Anche il Talmud era un ipertesto*, Roma, 27 ottobre 1997, intervista a Mediamente. Il testo integrale è disponibile all'indirizzo:
<http://www.mediamente.rai.it/home/bibliote/intervis/k/kolb.htm/>

R. Minuti, *Storiografia, riviste e reti: una transizione avviata*, in *Linguaggi e Siti: la Storia Online*, Firenze 6-7 Aprile 2000, Convegno SISSCO. Testo Minuti all'indirizzo <http://www.sissco.it/dossiers/internet/minuti-sem-apr-2000.html

R. Minuti, *Internet e il mestiere di storico. Riflessioni sulle incertezze di una mutazione*, in «Cromohs» 6 (2001), pp. 1-75. Articolo disponibile all'indirizzo: <URL: http://www.cromohs.unifi.it/6_2001/rminuti.html>

R. Minuti, *Le frontiere editoriali*, in G. Abbattista e A. Zorzi (a cura di), *Il documento immateriale*, *Ricerca storica e nuovi linguaggi*, La Storia – Consorzio italiano per le discipline storiche online. Disponibile all'indirizzo: <http://lastoria.unipv.it/dossier/minuti.htm>

P. Ortoleva, *Ipertesto*. Testo integrale all'indirizzo:
http://server.forcom.unito.it:8000/~studente/milanese/milanes3/peppino.html

P. Ortoleva, *L'argomentazione storica al tempo degli ipertesti*, in G. Abbatista e A. Zorzi (a cura di), *Il documento immateriale. Ricerca storica e nuovi linguaggi*. Testo disponibile all'indirizzo:
<http://lastoria.unipv.it/dossier/ortoleva.htm>

P. Ortoleva, *Società moderna e tecnologia*, Firenze, 21 ottobre 1997, intervista a Mediamente. Il testo integrale all'indirizzo:
<http://www.mediamente.rai.it/home/bibliote/intervis/o/ortoleva.htm>

L. Parolin, *Come cambia il concetto di autorità accademica con la rete*, in *Linguaggi e Siti: la Storia Online*, Firenze 6-7 Aprile 2000, Convegno SISSCO. <http://www.sissco.it/dossiers/internet/parolin-sem-apr-2000.html>

R. Di Quirico, *La rivoluzione informatica e le nuove frontiere dell'editoria accademica*. Testo integrale all'indirizzo:
<http://www.sissco.it/dossiers/internet/diquirico-sem-apr-2000.html>

Gino Roncaglia, *Ipertesti e argomentazione*, intervento presentato al convegno, *Le comunità virtuali e i saperi umanistici*, Milano, novembre 1997. Attulmente gli atti sono in via di pubblicazione a cura di P. Nerozzi, Bellman. Il testo integrale dell'intervento è disponibile all'indirizzo:
<http://www.merzweb.com/ipertesti/Ipertestieargomentazione.html>

M. Santoro, *Pubblicazioni cartacee e pubblicazioni digitali: quale futuro per la comunicazione scientifica?*, in Serge Noiret (a cura di), *Internet e il mestiere di Storico*. Il testo dell'intervento di Santoro all'indirizzo:
<http://www.sissco.it/dossiers/internet/santoro-sem-apr-2000.html>

A. Zorzi, *Le riviste tra due transizioni: crisi di ruolo e nuove pratiche editoriali*, in Medium-evo. Gli studi medievali e il mutamento digitale, I Workshop nazionale di studi medievali e cultura digitale, Firenze, 21-22 giugno 2001. Il testo integrale all'indirizzo:
<http://storia.unifi.it/_PIM/medium-evo/

A. Zorzi, Metafonti, in G. Abbatista e A. Zorzi (a cura di), Il *documento immateriale. Ricerca storica e nuovi linguaggi*. Testo disponibile all'indirizzo:
<http://lastoria.unipv.it/dossier/metafonti.htm>