

Progetto BibIt

Documento BibIt-1

Manuale di riferimento per la codifica testuale: livello 1

Fabio Ciotti

Revisione 3 – 11 marzo 2003

© Bibit

Introduzione

Questo manuale espone le norme per la codifica informatica di testi al livello 1 nell'ambito del progetto *Biblioteca Italiana* (BibIt). Il lavoro qui presentato va considerato come un riferimento di base per la creazione di testi in formato elettronico che entreranno a far parte della biblioteca digitale BibIt.

Il progetto Bibit ha scelto di adottare come linguaggio di markup il vasto schema di codifica sviluppato dalla *Text Encoding Initiative*, basato a sua volta sulla sintassi del linguaggio XML (*Extensible Markup Language*). La scelta di assumere questi standard come formati di riferimento del progetto è stata dettata da più considerazioni:

- completezza e qualità dello schema TEI:
- esigenza di uniformazione alle iniziative in corso a livello internazionale:
- necessità di garantire un adeguato livello di certificazione del testo:
- disponibilità di software adeguato per la creazione, archiviazione e diffusione dei dati:

XML è il sistema di codifica informatica dell'informazione testuale che presenta i migliori requisiti per le esigenze poste dalla ricerca scientifica sui testi. La prima caratteristica distintiva dello XML è quella di essere uno standard di pubblico dominio rilasciato ufficialmente dal W3 Consortium. XML è stato progettato appositamente per esser *indipendente* dai sistemi informativi, dai supporti e dispositivi digitali attualmente esistenti e futuribili. In sostanza le informazioni testuali codificate attraverso applicazioni XML godono della massima oggi possibile *portabilità* nello spazio e nel tempo.

Questo requisito è di estrema rilevanza in tutti i processi di gestione dell'informazione testuale e documentale che vedono coinvolti molteplici soggetti, eventualmente dislocati nello spazio e nel tempo, o che hanno l'obiettivo di diffondere le informazioni attraverso molteplici canali comunicativi informatizzati, quali quelli telematici, ma anche tradizionali come la stampa. Soprattutto esso fornisce le maggiori garanzie di preservazione dei dati in quei progetti di archiviazione digitale a lungo termine di dati sensibili, come è il caso del patrimonio culturale. I vantaggi offerti da XML, d'altra parte, non si limitano alla standardizzazione ed alla portabilità. Infatti, esso offre un modo sufficientemente semplice ma al contempo estremamente rigoroso e potente per rappresentare su supporto digitale fenomeni testuali di grande complessità, come quelli studiati dalla ricerca in ambito umanistico e letterario.

Per queste ragioni esso è stato assunto come linguaggio di base per lo schema di codifica realizzato dalla *Text Encoding Initiative* (TEI). Si tratta di un progetto internazionale che ha visto coinvolte le maggiori organizzazioni internazionali dedicate alla Informatica Umanistica (*Computing in Humanities*) al fine di definire uno schema di codifica per la realizzazione di testi elettronici con finalità di ricerca. La spinta che ha portato alla realizzazione di una simile iniziativa sono le stesse che abbiamo tracciato prima. L'esigenza di portabilità dei dati, infatti, è tanto più pressante quando

si tratta di trasferire su supporto digitale il patrimonio informativo di una biblioteca o di un archivio documentale. Questi progetti richiedono un notevole dispendio di risorse, che rischia di essere vanificato dalla rapida evoluzione tecnologica.

D'altra parte la tipologia di trattamenti testuali che vengono richiesti dalla ricerca umanistica sono molto più complessi rispetto a quelli della consueta gestione documentale in ambito commerciale e produttivo, e si dispongono su una articolata scala di astrazione. Basti citare a titolo di esempio i fenomeni studiati dalla critica testuale, a un estremo, e a quelli oggetto della narratologia, all'altro. Chiaramente uno schema di codifica per la ricerca umanistica deve essere in grado di rispondere a tutte queste esigenze, fornendo un *framework* comune in cui i diversi specialisti del testo possano riconoscere strumenti di lavoro adeguati, e nello stesso tempo mantenere l'aderenza alle esigenze del trattamento informatico e della standardizzazione dei formati di rappresentazione.

Data la vastità dello schema TEI, nell'ambito del progetto Bibit sono stati individuati diversi livelli di codifica ai quali un testo può essere sottoposto:

- livello 1: codifica della struttura editoriale del testo, di un limitato gruppo di fenomeni editoriali intralineari e linguistici;
- livello 2: codifica di una serie di fenomeni linguistici e semantici, dei riferimenti bibliografici, oltre alla eventuale introduzione di riferimenti incrociati;
- livello 3: codifica di fenomeni testuali complessi in vista di applicazione di analisi avanzate (struttura semantica, narrativa, retorica, morfosintattica, etc.);
- livello 4: trascrizione diplomatica di una fonte primaria;
- livello 5: edizione critica di un'opera.

Si noti che i livelli 3, 4 e 5 non vanno considerati come successivi su scala temporale o di complessità, ma piuttosto come livelli paralleli di articolazione del processo di codifica. È inoltre evidente che il livello 4 è applicabile esclusivamente alle fonti primarie.

Tutti i testi della collezione principale del progetto Bibit dovranno almeno presentare un livello di codifica 1. Questo documento illustra esclusivamente le norme di codifica richieste da tale livello.

Successivi documenti di riferimento saranno realizzati per specificare le norme di codifica per i livelli successivi. Per eventuali approfondimenti o per specifiche esigenze di codifica, inoltre, si rimanda alla documentazione originale del TEI, con l'avvertenza che essa si presenta di non facile comprensione per chi non possieda una certa dimestichezza con la sintassi XML, e con il linguaggio della documentazione tecnica di carattere informatico.

Principi editoriali generali

Le risorse testuali prodotte nel contesto del progetto BibIt – ad eccezione di edizioni diplomatiche di fonti primarie – vanno considerate in generale come vere e proprie *edizioni elettroniche*. Da questo punto di vista il rapporto con l'edizione cartacea fonte è circoscritto al solo contenuto testuale dell'opera nella forma linguistica e nella struttura editoriale con cui essa vi appare.

Di conseguenza il processo di digitalizzazione e codifica si limiterà esclusivamente a tale contenuto testuale, trascurando l'aspetto materiale della fonte e i materiali paratestuali la cui responsabilità intellettuale non è ascrivibile all'autore dell'opera (a meno che tali elementi non siano entrati a far parte dell'opera stessa nel corso della sua tradizione).

Dal punto di vista operativo questo significa che:

- tutti i materiali paratestuali non d'autore o comunque estranei al testo in sé, per come questo è attestato nella tradizione (frontespizio, introduzione, prefazione, indice dei contenuti, note, titoli correnti, numeri di pagina etc.), saranno omessi;

- le parole sillabate a fine riga saranno normalizzate;
- ogni fenomeno materiale non inerente al testo dell'opera occorrente sulle pagine sarà tralasciato;
- ogni fenomeno di evidenziazione sarà codificato in modo funzionale.

Per quanto riguarda il contenuto linguistico e ortografico del testo esso dovrà essere trascritto fedelmente nella forma in cui esso appare nella fonte, senza alcun intervento di normalizzazione. Il code set di riferimento per la rappresentazione dei caratteri è *Unicode* in codifica UTF-8. Dal punto di vista pratico questo significa che nel corso della codifica potranno essere usate tutte le lettere accentate e i segni diacritici presenti sulla tastiera. Per tutti gli altri caratteri e per gli alfabeti non latini potranno essere usati sia i caratteri Unicode sia le entità carattere definite negli insiemi pubblici rilasciati dalla ISO. Tuttavia se si inseriscono caratteri in alfabeti non latini è necessario utilizzare un font Unicode come Arial Unicode per ambiente Windows.

Fanno eccezione i caratteri sintattici di XML, per cui vanno sempre usate le entità secondo il seguente schema:

- **&**; : e commerciale (&)
- **<**; : minore (<)
- **>**; : maggiore (>)

Al posto dei caratteri della tastiera si possono utilizzare le entità anche per i seguenti segni diacritici, al fine di evitare il rischio che vada perduta la differenza tra il segno di apertura e quello di chiusura.

- **&lquo;**; : doppie virgolette aperte [“]
- **&rquo;**; : doppie virgolette chiuse [”]
- **&lquo;**; : virgoletta singola aperta [‘]
- **’**; : virgoletta singola chiusa [’]
- **&lquo;**; : doppie virgolette ad angolo (sergenti) aperte [«]
- **»**; : doppie virgolette ad angolo (sergenti) chiuse [»]
- **—**; : trattino lungo [—]
- **–**; : trattino medio [–]

Lo schema di codifica Biblt livello 1

Lo schema di codifica di seguito illustrato prevede la rappresentazione esplicita della macrostruttura del testo, di un limitato insieme di elementi intralineari, e una limitata documentazione del documento elettronico e della sua fonte. Esso fornisce gli elementi (e le relative indicazioni di uso) per rappresentare i seguenti generi testuali:

- testi in prosa (romanzi, raccolte di novelle, trattati)
- testi in versi (canzonieri, raccolte di liriche, poemi)
- testi drammatici (commedie, tragedie, sceneggiature)

I romanzi epistolari vanno trattati in generale come testi in prosa, anche se vengono forniti alcuni elementi per rappresentare eventuali caratteristiche peculiari di questo genere di testi. I dialoghi o le parti di opere che assumono la forma di un dialogo esplicito vanno trattati come testi drammatici. I prosimetri andranno codificati usando sia gli elementi per la prosa sia quelli per i versi. Per i casi

complessi di testi drammatici in versi, che daranno luogo a fenomeni di sovrapposizione tra struttura metrica e struttura editoriale, si dovrà ritenere prevalente la struttura editoriale e segmentare i versi tra le varie battute come indicato più avanti.

In linea generale tutti gli elementi e gli attributi descritti in questo documento sono da considerarsi obbligatori. Per quanto riguarda i valori degli attributi, laddove la DTD della TEI fornisce indicazioni formali queste sono state rispettate. Un valore che deve essere scelto entro un set di valori enumerati definiti nella DTD TEI, è definito *valore legale*. Negli altri casi viene fornito un elenco di valori standard.

Infine, come norma generale, e in assenza di eventuali software di editing XML specifici, si raccomanda di seguire le seguenti direttive:

- evitare qualsiasi minimizzazione dei tag nella redazione dei documenti. Ciò significa che tutti gli elementi dovranno avere entrambi i tag (eccezion fatta ovviamente per gli elementi vuoti) e che i valori di attributo dovranno essere racchiusi tra doppi o singoli apici;
- gli elementi vuoti devono avere la seguente sintassi `<NomeElemento/>`;
- i nomi degli elementi e degli attributi sono sensibili alla differenza maiuscolo/minuscolo e dunque è necessario che i tag siano inseriti esattamente nella forma in cui sono esposti in questo manuale;
- il valore degli attributi deve essere sempre inserito tra apici doppi o singoli. Per motivi di comodità si richiede di usare gli apici singoli.

XML e DTD declaration

La prima istruzione che si trova in un documento XML è la *XML declaration* che ha la seguente forma

```
<?xml version='1.0' encoding='UTF-8' ?>
```

Subito dopo andrà inserita la *document type declaration* che ha il fine di dichiarare a quale DTD esso si riferisca:

I documenti conformi al primo livello di codifica TIL, il cui insieme di elementi è qui descritto, sono conformi alla DTD TEI. Per la precisione la DTD utilizzata è composta dal *core tag set*, dai *base tag set mixed* (che permette di utilizzare molteplici *base set* contemporaneamente), *prose*, *verse* e *drama* e dagli *additional tag set figure* (di cui tuttavia si usano solo pochi elementi) *transcr* e *linking*. Sono state effettuate anche alcune estensioni allo schema standard seguendo le direttive specificate nel capitolo 29 «Modifying and Customizing the TEI DTD» delle *Guidelines*:

- soppressione di alcuni elementi (cfr. infra)
- inserimento di un elemento `<ps>` nella classe *x.divbot* con content model *paraContent*
- aggiunta dell'elemento `<space>` nella classe *x.common*
- aggiunta dell'attributo `type` all'elemento `<space>`

Questa è la DTD Declaration per personalizzare la DTD TEI.2:

```
<!DOCTYPE TEI.2 PUBLIC "-//TEI P4//DTD Main Document Type//EN" "tei2.dtd"
[
<!ENTITY % TEI.XML 'INCLUDE' >
<!ENTITY % TEI.mixed 'INCLUDE' >
<!ENTITY % TEI.prose 'INCLUDE' >
<!ENTITY % TEI.drama 'INCLUDE' >
```

```
<!ENTITY % TEI.verse 'INCLUDE' >
<!ENTITY % TEI.linking 'INCLUDE' >
<!ENTITY % TEI.figures 'INCLUDE' >
<!ENTITY % TEI.transcr 'INCLUDE' >
<!ENTITY % TEI.extensions.ent SYSTEM 'bibit-l1Extns.ent' >
<!ENTITY % TEI.extensions.dtd SYSTEM 'bibit-l1Extns.dtd' >
]>
```

Il file ‘bibit-l1Extns.ent’ contiene le seguenti dichiarazioni:

```
<!-- I seguenti elementi sono eliminati -->
<!ENTITY % alt "IGNORE">
<!ENTITY % altGrp "IGNORE">
<!ENTITY % analytic "IGNORE">
<!ENTITY % biblFull "IGNORE">
<!ENTITY % biblScope "IGNORE">
<!ENTITY % biblStruct "IGNORE">
<!ENTITY % camera "IGNORE">
<!ENTITY % caption "IGNORE">
<!ENTITY % dateRange "IGNORE">
<!ENTITY % dateStruct "IGNORE">
<!ENTITY % divGen "IGNORE">
<!ENTITY % formula "IGNORE">
<!ENTITY % index "IGNORE">
<!ENTITY % join "IGNORE">
<!ENTITY % joinGrp "IGNORE">
<!ENTITY % lg1 "IGNORE">
<!ENTITY % lg2 "IGNORE">
<!ENTITY % lg3 "IGNORE">
<!ENTITY % lg4 "IGNORE">
<!ENTITY % lg5 "IGNORE">
<!ENTITY % link "IGNORE">
<!ENTITY % linkGrp "IGNORE">
<!ENTITY % listBibl "IGNORE">
<!ENTITY % meeting "IGNORE">
<!ENTITY % monogr "IGNORE">
<!ENTITY % move "IGNORE">
<!ENTITY % postBox "IGNORE">
<!ENTITY % postCode "IGNORE">
<!ENTITY % recordingStmnt "IGNORE">
<!ENTITY % recording "IGNORE">
<!ENTITY % equipment "IGNORE">
<!ENTITY % broadcast "IGNORE">
```

```

<!ENTITY % setting "IGNORE">
<!ENTITY % sound "IGNORE">
<!ENTITY % street "IGNORE">
<!ENTITY % tech "IGNORE">
<!ENTITY % teiCorpus.2 "IGNORE">
<!ENTITY % timeRange "IGNORE">
<!ENTITY % timeline "IGNORE">
<!ENTITY % view "IGNORE">
<!ENTITY % when "IGNORE">
<!-- il seguente elemento è ridefinito -->
<!ENTITY % space "IGNORE">
<!-- le seguenti classi sono estese -->
<!--
        modifica classe divbot con aggiunta elemento ps
        ps può comparire alla fine di una div*
-->
<!ENTITY % x.divbot "ps|" >
<!--
        modifica classe common con aggiunta elemento space
        space può comparire anche tra elementi chunk come p
-->
<!ENTITY % x.common "space|" >

```

Il file ‘bibit-11Extns.dtd’ contiene le seguenti dichiarazioni:

```

<!-- ridefinizione elemento space: aggiunto attributo type
        Scopo: usare space con type per segnalare segmentazioni dei capitoli non
        identificate chiaramente da intertitoli etc.
-->
<!ELEMENT space %om.RO; EMPTY>
<!ATTLIST space %a.global;
dim (horizontal | vertical) #IMPLIED
extent CDATA #IMPLIED
resp CDATA #IMPLIED
type CDATA #IMPLIED
TEIform CDATA 'space' >
<!-- dichiarazione elemento ps per codificare postscriptum
        classe: divbot
        Content model: paraContent
-->
<!ELEMENT ps %om.RO; %paraContent;>
<!ATTLIST p
        %a.global;>

```

Tali modifiche ai tag set standard TEI sono tali da configurare una DTD autonoma, individuata da un identificatore pubblico formale (*Public Identifier*):

```
-//BIBIT//DTD BibIt L1 TEI Document Type//EN
```

Infine si ricorda che in XML è obbligatorio indicare un system identifier per la DTD. Di conseguenza la DTD Declaration per i testi del progetto sarà:

```
<!DOCTYPE TEI.2 PUBLIC "-//BIBIT//DTD BibIt L1 TEI Document Type//EN" "bibit-11.dtd">
```

Struttura di un documento TEI

L'elemento radice per i documenti TEI è <TEI.2>. Ogni documento TEI, poi, è diviso in due parti, individuate rispettivamente dagli elementi <teiHeader> e <text>, entrambi obbligatori.

La prima parte comprende le informazioni (metadati) di carattere bibliografico ed editoriale relative al documento TEI e alla sua fonte. Si tratta di informazioni bibliografiche (sia per il testo fonte sia per il testo MRF), informazioni relative allo stato di revisione del testo MRF, ai principi di codifica e di trascrizione seguiti. L'elemento <text> contiene la versione elettronica del testo vero e proprio.

La struttura generale di ogni documento è dunque la seguente:

```
<TEI.2>
    <teiHeader>
        ...
        [metadati]
        ...
    </teiHeader>
    <text>
        ...
        [contenuto testuale]
        ...
    </text>
</TEI.2>
```

Il TEI Header e la documentazione del testo elettronico

La documentazione del testo elettronico è un aspetto che spesso viene posto in secondo ordine nelle discussioni dedicate al tema della codifica testuale. Essa invece è un requisito fondamentale di uno schema di codifica usato in applicazioni di tipo scientifico: sia al fine di permettere una adeguata catalogazione del testo elettronico all'interno di archivi o *corpora* disponibili su stazioni di lavoro locali o in rete; sia per facilitare una corretta interpretazione della codifica da parte di utenti distanti nello spazio e nel tempo dal responsabile primario della stessa.

Una corretta documentazione del testo elettronico dovrebbe provvedere almeno le seguenti informazioni:

- individuazione del testo elettronico attraverso le sue determinazioni bibliografiche: titolo, autore, luogo e data di edizione, etc.;
- certificazione della responsabilità editoriale del testo, anche nel caso in cui la codifica ha avuto diversi responsabili;
- indicazione della fonte su supporto originale del testo per i testi elettronici che sono edizioni di testi già editi su stampa o trascrizioni di manoscritti.;

- documentazione accurata delle metodologie di rappresentazione dei vari fenomeni testuali, delle scelte di approcci disciplinari o teorici che permettano di interpretare correttamente i simboli usati nella codifica del testo, delle eventuali correzioni e modifiche introdotte rispetto alla fonte.;

Nello schema di codifica TEI l'elemento `<teiHeader>` svolge la funzione di raccogliere ed organizzare tutti i metadati relativi al documento codificato ed è a sua volta è diviso in varie sezioni. Per quanto riguarda i testi al livello di codifica 1 la creazione del `<teiHeader>` va fatta usando il seguente modello, dove le parti tra parentesi quadre vanno sostituite da parte del codificatore con le relative informazioni:

```
<teiHeader>
<fileDesc>
<titleStmt>
  <!-- qui va il titolo del testo fonte -->
  <title>[titolo del testo]</title>
  <!-- qui vanno cognome e nome dell'autore -->
  <author>[Nome autore]</author>
  <respStmt>
    <resp>Codifica</resp>
    <name>[nome codificatore]</name>
  </respStmt>
</titleStmt>
<publicationStmt>
  <publisher>BibIt</publisher>
  <pubPlace>Roma</pubPlace>
  <date>2002</date>
</publicationStmt>
<seriesStmt>
  <title>Collezione universale BibIt</title>
</seriesStmt>
<!-- qui vanno le indicazioni bibliografiche dettagliate relative alla fonte -->
<sourceDesc>
  <bibl>
<!-- qui va il titolo ed eventuali sottotitoli del testo fonte come appaiono nel frontespizio -->
    <title>[titolo del testo]</title>
    <author>[nome autore]</author>
    <!-- qui vanno i nomi degli eventuali curatori del testo fonte -->
    <editor>[nome del curatore]</editor>
    <publisher>[editore del testo fonte]</publisher>
    <pubPlace>[luogo di edizione del testo fonte]</pubPlace>
    <date>[data di edizione del testo fonte]</date>
  </bibl>
```



```

</sourceDesc>
</fileDesc>
<profileDesc>
    <!-- qui vanno specificate le lingue presenti nel testo -->
    <langUsage>
        <language id='codice'[lingua]</language>
    </langUsage>
</profileDesc>
<!-- qui vanno documentati gli interventi effettuati sul documento elettronico -
->
    <revisionDesc>
        <change>
            <date>[data]</date>
            <respStmt>
                <name>[nome codificatore]</name>
            </respStmt>
            <item>Codifica livello 1</item>
        </change>
    </revisionDesc>
</teiHeader>

```

Elementi comuni a ogni testo

In questo paragrafo vengono introdotti gli elementi necessari a rappresentare la macrostruttura comune a ogni genere testuale e le sue suddivisioni strutturali interne.

La macrostruttura del testo

Ogni testo deve iniziare con l'elemento obbligatorio **<text>**, che segue immediatamente il **<teiHeader>**, e segnala l'inizio del contenuto testuale vero e proprio.

Questo elemento si articola a sua volta in tre elementi di livello inferiore, che riflettono la divisione tra materiali peritestuali e materiali testuali veri e propri, per come questa si è venuta formando nella tradizione della stampa in occidente:

<front>: contiene tutti i materiali di tipo avantestuale che introducono un testo nelle edizioni a stampa, dalla pagina del titolo, al frontespizio, ad eventuali introduzioni o prefazioni.

<body>: contiene il testo vero e proprio, il suo corpo.

<back>: contiene tutti i materiali peritestuali che possono essere rinvenuti nelle pagine finali di un testo stampato, quali postfazioni, glossari, indici, etc.

Riassumendo la macrostruttura di un testo unitario è la seguente:

```

<text>
    <front>
        [materiali avantestuali]
    </front>
    <body>
        [testo]

```

```
</body>
<back>
[materiali che seguono il testo]
</back>
</text>
```

Le divisioni del testo

Un testo ed il suo peritesto editoriale si presentano in genere come una sequenza di espressioni linguistiche, dotata generalmente di una articolazione interna. Tale articolazione suddivide e raggruppa (in base a principi di coerenza logico-semantica e di posizione funzionale nella struttura complessiva) le espressioni linguistiche in sezioni testuali, e definisce una struttura gerarchica a più livelli, che può essere più o meno profonda, e in cui ogni singolo livello può presentare una o più istanze. Le sezioni testuali sono di norma evidenziate dalla forma materiale del documento originale, mediante spaziature, pagine bianche, stili di carattere, etc.

Tradizionalmente tali suddivisioni strutturali hanno assunto dei nomi specifici in base al genere cui il testo in oggetto appartiene: “capitolo”, “paragrafo”, “sezione”, “parte”, per la prosa, “poema”, “canto”, per i versi, “atto”, “scena”, per la scrittura drammaturgica. A loro volta ogni livello può occorrere una o più volte, avendo così anche un indice ordinale: “primo capitolo”, “secondo capitolo”, “prima sezione”, “seconda sezione”, e così via.

La codifica di questa struttura nella TEI può essere attuata in diversi modi. Quello qui proposto si basa su una serie di elementi indicizzati in base al livello gerarchico a cui si riferiscono: **<div1>**, **<div2>**, **<div3>** ... **<div7>**. Attraverso questo sistema si possono codificare fino a sette livelli di profondità, ampiamente sufficienti per rappresentare la struttura di testi letterari. Tutti questi elementi sono dotati di un attributo obbligatorio **type**, con il quale si deve specificare il tipo di divisione in base a quanto esplicitamente espresso nel documento originale o ad una attribuzione effettua da chi effettua la codifica. Nei paragrafi successivi dedicati ad ogni genere sarà fornita la lista di valori da utilizzare per le più comuni partizioni testuali.

Un testo in prosa codificato avrà dunque la seguente forma:

```
<body>
<div1 type='capitolo'>
  [qui vanno i paragrafi del testo]
</div1>
<div1 type='capitolo'>
  [qui vanno i paragrafi del testo]
</div1>
</body>
```

Elementi comuni a tutte le divisioni

Alcune strutture testuali possono apparire indifferentemente nel contesto di divisioni di ogni livello strutturale. Per codificare tali componenti sono previsti degli elementi specifici. Il primo di essi è l'elemento per la codifica degli intertitoli:

<head>: contiene il titolo di una suddivisione strutturale nel testo e nel peritesto.

Questo elemento è obbligatorio e va inserito anche se le titolazioni assumono forme puramente convenzionali come “capitolo I” o “I”.

Ad esempio:

```
<text>
```

```
<div1 type='capitolo'><head >La casa e la talpa</head>
[testo]
</div1>
```

Nel caso di intertitoli complessi, che possono apparire o meno su righe diverse vanno usati più elementi **<head>**.

L'elemento **<argument>** va usato per codificare le brevi descrizioni del contenuto di una sezione che seguono il titolo, tipiche in testi di epoca medievale e moderna. Ad esempio:

```
<div1 type='novella'>
<head >Novella prima</head>
<argument><p>Ser Cepparello con una falsa confessione inganna ...
</p></argument>
...
</div1>
```

Si noti che il contenuto di un elemento **<argument>** deve essere a sua volta opportunamente codificato mediante elementi di livello blocco (**<p>**, **<lg>**, **<l>**).

L'elemento **<epigraph>** contiene le epigrafi che possono occorrere a ogni livello nel testo e nel peritesto, in testa a una determinata sezione testuale (sia in prosa sia in versi) prima o dopo il titolo della stessa.

Qualora una epigrafe sia inserita in una pagina autonoma o non possa essere ricondotta direttamente a una particolare sezione testuale essa va codificata mediante un elemento **<div1>**, **<div2>**, etc. dotato dell'attributo **type=epigrafe**.

Il contenuto di una epigrafe va codificato con gli elementi **<p>** se il testo è in prosa, mentre epigrafi in versi vanno ulteriormente codificate mediante gli opportuni elementi per la rappresentazione di testi poetici. L'eventuale riferimento bibliografico, più o meno completo, va codificato mediante l'elemento **<bib1>** che conterrà l'indicazione bibliografica.

Se il testo presenta formule di commiato a fine capitolo o alla conclusione dell'opera esse vanno codificate mediante l'elemento **<trailer>**.

Le strutture peritestuali

Come anticipato sopra tutti i materiali peritestuali che non sono riconducibili alla responsabilità diretta dell'autore vanno omessi, a meno che non siano divenuti parti integranti del testo nel corso della sua tradizione. Dunque la pagina del titolo o frontespizio, il retrofrontespizio, il colophon, le introduzioni, le postfazioni, gli indici e glossari e ogni altra pagina o componente dell'edizione fonte che non siano direttamente parte dell'opera non saranno riprodotte nell'edizione elettronica.

Se invece esistono sezioni testuali d'autore (introduzioni, dediche, epigrafi, etc.) esse vanno digitalizzate e codificate opportunamente. Si noti che talvolta sezioni testuali che sono intitolate con formule tipiche di sezioni peritestuali sono da considerare in effetti parti del testo. Ad esempio la "Introduzione" manzoniana ai *Promessi Sposi* è parte dell'artificio narrativo e va dunque inserita all'interno dell'elemento **<body>**.

Nei casi in cui i materiali peritestuali vadano trascritti, le eventuali sezioni presenti all'interno del peritesto debbono essere codificate mediante l'elemento **<div1>** (seguito da elementi **<div2>** etc., per eventuali sottodivisioni), mentre titoli ed epigrafi avranno il trattamento indicato nel paragrafo precedente. L'attributo **type** dovrà essere utilizzato per indicare il tipo di sezione. La scelta del suo valore **type** va effettuata in base alla nomenclatura del documento di origine normalizzata secondo il seguente elenco:

- **introduzione**
- **prefazione**
- **dedica**
- **epigrafe**
- **appendice**
- **glossario**
- **indice**
- **postfazione**
- **sommario**

Il contenuto di queste partizioni può essere ovviamente testo in prosa o in versi (con eventuali ulteriori partizioni strutturali), nel caso di introduzioni, dediche, pre o postfazioni; ovvero liste ordinate, nel caso di indici di luoghi o glossari. Una introduzione generale al testo sarà codificata nel seguente modo:

```
<front>
<div1 type='intr'>
[qui va il testo dell'introduzione]
</div1>
</front>
```

Mentre un glossario a fine testo sarà codificato come segue:

```
<back>
<div1 type='glos'>
[qui va il glossario]
</back>
```

Sono inoltre disponibili gli elementi `<prologue>` e `<epilogue>` per rappresentare prologo o epilogo quando questi siano nettamente distinti dal testo principale dell'opera. In caso contrario essi vanno inseriti all'interno del `<body>` segnalati da elementi `<div1>` dotati di attributo `type` con valori `prologo` e `epilogo`

La codifica di testi in prosa

I testi in prosa sono in genere costituiti da una serie di sezioni, talvolta articolate in ulteriori sottosezioni, composte a loro volta da una sequenza di blocchi testuali denominati in italiano capoversi ed in inglese *paragraph*. Talvolta all'interno dei capoversi possono occorrere blocchi di testo strutturati in forma di liste, note e blocchi di testo citato (la cui responsabilità è altra da quella dell'autore del testo).

Per quanto riguarda la tipologia delle divisioni strutturali, che come noto vanno registrate mediante l'attributo `type` dei vari elementi `<div#>`, vanno utilizzate le seguenti sigle:

- **sezione**
- **parte**
- **capitolo**
- **paragrafo**
- **dialogo**
- **epistola**

- **discorso**
- **orazione**
- **predica**
- **satira**
- **sermone**

Un trattamento particolare va riservato alle epistole, che sono dotate di una serie di costrutti peculiari per i quali sono previsti alcuni elementi speciali.

L'elemento paragrafo

L'elemento fondamentale per la codifica dei testi in prosa è <p>. Esso rappresenta la più piccola unità testuale di tipo strutturale in cui può essere diviso un testo in prosa (escludendo le unità linguistiche), ed è rappresentato in quasi tutti i testi a stampa di epoca moderna attraverso vari dispositivi presentazionali (indentazioni, salti di linea, etc.).

Al suo interno possono occorrere caratteri validi, tutti gli elementi di blocco e intralinee descritti nel seguito, gli elementi per i rinvii ipertestuali e gli elementi per la codifica di brani in versi.

L'elemento <p> va ovviamente utilizzato non solo all'interno del testo vero e proprio, ma ovunque appaia del testo in prosa diviso in blocchi distinti, poiché segnala la più piccola unità strutturale di un testo.

Note

L'elemento <note> va utilizzato per la codifica di note e annotazioni del testo. Esso in genere può essere impiegato sia per segnalare note dovute all'autore di un'opera, sia note presenti nel testo originale aggiunte da curatori o traduttori, sia, infine, note editoriali aggiunte dal curatore della trascrizione elettronica. Rammentiamo tuttavia che per il livello di codifica 1 delle fonti secondarie tutti i materiali di apparato presenti eventualmente sull'originale vanno omessi e che *solo le note di autore* debbono essere registrate.

La scelta circa il luogo del documento in cui posizionare la nota è complicata. Dopo aver considerato le varie opzioni si è optato per inserire il testo di una nota direttamente nel luogo in cui nel testo a stampa appare il rimando, ovviamente opportunamente racchiuso dall'elemento.

L'elemento <note> va qualificato mediante i seguenti attributi:

n: indica il simbolo grafico o il numero utilizzato per segnalare il rimando alla nota nel testo.

resp: indica il responsabile della nota. Il valore da attribuire alle note d'autore è

- **aut:** nota dovuta all'autore del testo

place: indica la posizione della nota nel testo originale. I valori possibili sono:

- **foot:** nota a piè di pagina
- **end:** nota a fine capitolo
- **sin:** nota al margine sinistro
- **des:** nota al margine destro
- **inter:** nota interlineare

La codifica di lettere ed epistole

Le lettere ed epistole possono occorrere in vari luoghi di un documento, sia nel peritesto sia all'interno del testo. Un caso particolare è costituito dal romanzo epistolare.

In generale una lettera va trattata come una divisione testuale qualificata dall'attributo **type=ep**. Tuttavia, essa può presentare una serie di strutture sia in apertura sia in chiusura. I seguenti elementi opzionali possono essere utilizzati per registrare tali strutture:

<dateline>: contiene le indicazioni di tempo e luogo di una lettera o di una dedica

<byline>: contiene le indicazioni di responsabilità di una lettera

<salute>: contiene le formule di saluto e le dediche

<signed>: contiene la firma dell'autore in calce a una epistola

<opener>: raggruppa i precedenti elementi quando essi appaiono in testa alla lettera

<closer>: raggruppa i precedenti elementi quando essi appaiono in testa alla lettera

Questi elementi possono essere usati anche per rappresentare elementi testuali come dediche o ringraziamenti che possono comparire in diversi luoghi del testo, anche in sezioni paratestuali.

Blocchi testuali non determinati

In alcuni nei testi in prosa si incontrano elementi testuali di livello strutturale simile al paragrafo ma che non hanno la medesima valenza semantica di un paragrafo (o di componenti testuali come specializzati come argomenti o epigrafi).

Questo tipo di strutture testuali può essere codificato mediante l'elemento **<ab>**, cui va associato un attributo **type** che segnala il presumibile ruolo funzionale assunto dal blocco in questione. Il valore dell'attributo va scelto in modo opportuno al fine di rendere chiaro quale sia tale ruolo. Se è necessario segnalare anche l'aspetto grafico di tali elementi va usato l'attributo globale **rend**.

L'elemento **<ab>** può essere utilizzato anche in contesti in cui non sia del tutto chiaro se si ha a che fare con segmenti testuali in prosa o in versi, soprattutto nel contesto di testi drammatici.

La codifica dei testi in versi

I testi in versi possono essere raccolte di poesie, talvolta divise in parti, o poemi, di norma divisi in canti. In entrambi i casi le suddivisioni strutturali del testo vanno codificate mediante elementi **<div>** qualificati opportunamente con l'attributo **type** in base al seguente schema:

- **canto**: contiene un canto nell'ambito di un poema
- **parte**: se la raccolta è divisa in parti dotate o meno di titoli
- **poesia**: contiene una singola poesia nell'ambito di una raccolta

I componimenti poetici sono a loro volta composti da una sequenza di unità di base dette versi che assumono diverse forme metriche le quali determinano la distribuzione dei versi in gruppi detti strofe.

Ogni componimento poetico e ogni sua partizione strofica interna vanno codificati mediante l'uso di elementi **<lg>** opportunamente annidati per documentare schemi strofici a più livelli strutturali. Per qualificare il tipo di componimento poetico occorre associare un attributo **type** all'elemento **<lg>** di livello più alto. Non vanno invece indicate esplicitamente le tipologie strofiche. I valori di tale attributo **type** vanno scelti dal seguente elenco:

1. poemi

- **ottava**
- **terzina**
- **endecasillabo-sciolto**

2. forme regolari della poesia lirica

- ballata
- barzelletta
- canzonetta melica
- ballata romantica
- lauda
- canzone
- aria o arietta
- canzone ode
- canzonetta
- ode-pindarica
- sestina
- stanza
- madrigale
- rondò
- sonetto
- strambotto
- rispetto
- stornello
- villanella
- villotta

3. forme regolari della poesia discorsiva

- endecasillabo-sciolto
- lassa
- ottava
- serventese
- terza-rima
- capitolo-bucolico
- capitolo-elegiaco
- capitolo-ternario
- egloga
- elegia

4. forme libere della metrica tradizionale

- frottola
- mottetto
- poesia-drammatica

5. discorso libero in endecasillabi e settenari

- canzone-libera

- madrigale
- poesia-drammatica

6. metrica barbara

- strofa-saffica
- strofa-alcaica
- strofa-asclepiadea

7. metrica libera

- verso-libero

I singoli versi invece vanno codificati mediante l'elemento <1>. Qualora si intenda rendere (nei limiti della possibilità) la disposizione originale del testo poetico si può utilizzare l'attributo globale **rend**.

In questo modo, un singolo sonetto all'interno di una raccolta sarà in primo luogo incluso in un elemento <div1 type='poesia'>, e di seguito in un elemento <lg type='sonetto'>; le varie sezioni del sonetto andranno a loro volta codificate mediante elementi <lg>:

```
<div1 type='poesia'><head>LXXXVI</head>
<lg type='sonetto'>
<lg>
<l>S'i' fosse foco, arderéi 'l mondo;</l>
<l>s' i' fosse vento, lo tempesterei;</l>
<l>s'i' fosse acqua, i' l'annegherei;</l>
<l>s'i' fosse Dio, mandereil'en profondo;</l>
</lg>
<lg>
<l>s'i' fosse papa, sare' allor giocondo,</l>
<l>ché tutti cristiani imbrigherei;</l>
<l>s'i' fosse 'mperator, sa' che farei?</l>
<l>A tutti mozzarei lo capo a tondo.</l>
</lg>
<lg>
<l>S'i fosse morte, andarei da mio padre;</l>
<l>s'i' fosse vita, fuggirei da lui:</l>
<l>similmente farìa da mi' madre.</l>
</lg>
<lg>
<l>S'i' fosse Cecco, com'i' sono e fui,</l>
<l>torrei le donne giovani e leggiadre:</l>
<l>e vecchie e laide lasserei altrui.</l>
</lg>
</lg>
</div1>
```


Le prime terzina dell'inferno, invece, andranno codificate come segue:

```
<div1 type='parte'>
<head>Inferno</head>
<div2 type='canto'>
<head>Canto I</head>
<lg type='terzina'>
<l>Nel mezzo del cammin di nostra vita</l>
<l>mi ritrovai per una selva oscura,</l>
<l>ché la diritta via era smarrita.</l>
</lg>
<lg>
<l>Ahi quanto a dir qual era è cosa dura</l>
<l>esta selva selvaggia e aspra e forte</l>
<l>che nel pensier rinova la paura!</l>
</lg>
<lg>
<l>Tant'è amara che poco è più morte;</l>
<l>ma per trattar del ben ch'i' vi trovai,</l>
<l>dirò de l'altre cose ch'i' v'ho scorte.</l>
</lg>
```

La codifica dei testi drammatici

I testi drammatici oltre a presentare una particolare struttura interna, richiedono anche un trattamento speciale per i materiali peritestuali. Infatti tali testi possono anteporre al dramma vero e proprio una serie di informazioni relative al modo in cui esso va messo in scena.

Per questa ragion il **<front>** di un testo drammatico include alcuni elementi specializzati, che potranno essere utilizzati per codificare tale genere di informazioni. L'elemento **<set>** contiene la descrizione del luogo e del tempo in cui si collocano gli eventi del dramma. L'elemento **<castList>** contiene l'elenco delle *dramatis personae*. Esso è costituito da una serie di elementi **<castItem>**. Questi a loro volta contengono l'elemento obbligatorio **<role>**, all'interno del quale va inserito il nome del personaggio. Se è presente anche una breve descrizione del personaggio essa va inserita all'interno dell'elemento **<roledesc>**. Si veda il seguente esempio, tratto dalla *Mandragola* di Machiavelli:

```
<front>
...
<castlist>
<head>PERSONAGGI</head>
<castItem><role>Callimaco</role></castItem>
<castItem><role>Siro</role></castItem>
<castItem><role>Messer Nicia</role></castItem>
<castItem><role>Ligurio</role></castItem>
<castItem><role>Sostrata</role></castItem>
<castItem><role>Frate Timoteo</role></castItem>
```

```

<castItem><role>Una donna</role></castItem>
<castItem><role>Lucrezia</role></castItem>
...
</front>

```

Due altri componenti tipici di un testo drammatico possono comparire nell'ambito delle sezioni peristestuali: il prologo e l'epilogo. Per codificare tali componenti sono disponibili due elementi **<prologue>** e **<epilogue>**. Il contenuto sarà poi codificato in accordo alla sua natura testuale. Ad esempio nella *Mandragola* esso è una canzone:

```

<front>
[...]
<prologue>
<head>PROLOGO</head>
<lg type="canzone">
<l>Iddio vi salvi, benigni uditori,</l>
<l>quando e' par che dependa</l>
<l>questa benignità da lo esser grato.</l>
...
</lg>
</prologue>
[...]
```

Parallelamente ai testi in prosa e in versi, anche i testi drammatici presentano un articolazione in sezioni gerarchiche. Tali sezioni vanno codificate mediante elementi **<div>**, qualificati dall'attributo **type**. I valori specifici per questo genere di testi che possono essere assegnati a tale attributo sono:

- **atto**
- **scena**

La struttura interna di un testo drammatico, invece, è di norma caratterizzata da tre componenti: nome del personaggio che parla, battuta e didascalie. La codifica di queste struttura va effettuata mediante i seguenti elementi:

<sp>: racchiude ogni singola battuta, incluso il nome del personaggio

<speaker>: contiene il nome del personaggio che parla.

<stage>: contiene le indicazioni di scena e le didascalie

Ad esempio, sempre dalla *Mandragola*:

```

<div1 type='atto'>
<head>ATTO PRIMO</head>
<div2 type='scena'>
<head>SCENA PRIMA</head>
<stage>Callimaco, Siro.</stage>
<sp><speaker>CALLIMACO</speaker>
<p>Siro, non ti partire, i' ti voglio un poco.</p></sp>
<sp><speaker>SIRO</speaker>

```

```
<p>Eccomi .</p></sp>
```

```
...
```

```
</div2>
```

Il testo del discorso pronunciato dal personaggio vero e proprio ovviamente va codificato opportunamente in base alla sua natura prosastica o poetica. Nel primo caso esso sarà pertanto inserito all'interno di un elemento `<p>`. Nel secondo caso esso sarà codificato mediante gli elementi `<lg>` e `<l>`. Un problema in cui si può incorrere in questo secondo caso è quella della sovrapposizione gerarchica tra la struttura drammatica e quella strofica: ad esempio un verso potrebbe essere suddiviso tra due battute. Purtroppo non esiste una soluzione generale a questo genere di problemi, che in ambito XML rientrano nella classe delle *overlapping hierarchies*.

Una soluzione che si può adottare per i documenti tei è l'adozione dell'attributo `part` sull'elemento `<l>`. Il suo valore indica se il verso è completo o meno e di quale parte si tratti. I valori da usare per questo attributo sono:

x: il verso è metricamente incompleto

i: parte iniziale di un verso incompleto

m: parte intermedia di un verso incompleto

f: parte finale di un verso incompleto

Ad esempio:

```
<sp><speaker>[Personaggio]</speaker>
```

```
<l part='I'>Testo poetico [...]</l>
```

```
</sp>
```

```
<sp><speaker>[Personaggio]</speaker>
```

```
<l part='F'>Testo poetico[...]</l>
```

```
</sp>
```

Citazioni

Se il testo presenta citazioni di altri testi intralineari o evidenziate in modo esplicito da spazi o cambiamento della dimensione del carattere questi vanno codificati mediante l'elemento `<quote>`. Se la citazione è in un blocco graficamente autonomo va specificato l'attributo `rend='block'`. Il testo citato va codificato opportunamente con gli elementi per i blocchi in verso o in prosa. Esempi:

```
<quote><p>Testo [...]</p></quote>
```

```
<quote rend='block'>
```

```
  <lg>
```

```
    <l>Testo poetico[...]</l>
```

```
  </lg>
```

```
</quote>
```

Qualora il testo citato sia completo, ad esempio una lettera, esso va codificato con un elemento `<text>`, seguito da `<body>`, e dalle successive `<div>`. Ad esempio:

```
<quote rend='block'>
```

```
  <text>
```

```
<body>
<opener><dateline>data</dateline></opener>
<head></head>
<p>Testo [...]</p>
<closer><signed>Firma</signed></closer>
</body>
</text>
```

</quote>

Elementi intralineari

Gli elementi intralineari nella TEI sono tutti quegli elementi che possono comparire all'interno dei vari elementi strutturali e che di norma si applicano a sequenze di parole o frasi. Tra questi rientrano i vari fenomeni di evidenziazione, i nomi di persona e luogo, le date.

Evidenziazioni

Per evidenziazioni si intende quella serie di artifici tipografici utilizzati per mettere in rilievo un certo segmento testuale rispetto al contesto e richiamare l'attenzione del lettore. Nei testi a stampa esse consistono di norma in cambiamenti di stile del font (corsivo, grassetto, maiuscoletto, etc.), mentre nei manoscritti possono consistere in sottolineature o spaziature.

Le evidenziazioni di norma svolgono diverse funzioni strutturali: il corsivo ad esempio può essere usato per segnalare una enfasi, un termine tecnico o in lingua straniera, un titolo di monografia. Tali fenomeni vanno codificati esplicitando, per quanto possibile, il loro ruolo funzionale piuttosto che segnalando la loro apparenza fisica. A tale fine possono essere utilizzati i seguenti elementi.

Le espressioni che vengono enfatizzate per motivi stilistici o linguistici vanno codificate mediante l'elemento **<emph>**.

I termini in lingua straniera (esclusi i termini dialettali) dovranno essere codificati mediante l'elemento **<foreign>**. Si noti che è necessario segnalare tutti termini in lingue straniere anche se essi non sono esplicitamente segnalati sulla fonte. L'attributo **lang** servirà ad indicare la lingua in oggetto: il suo valore deve essere espresso usando le sigle ISO 639. Si ricordi che le lingue presenti in un testo debbono essere adeguatamente documentate nella sezione **<langUsage>** del **<teiHeader>**.

Nel caso delle evidenziazioni associate ai titoli in citazioni bibliografiche, si deve utilizzare il tag **<title>**.

Qualora non si intenda o non si possa collegare il fenomeno grafico con una funzione testuale ben determinata si deve utilizzare l'elemento **<hi>**, che segnala la pura presenza di un fenomeno grafico. In questo caso il tag va obbligatoriamente determinato con l'attributo **rend**.

L'attributo **rend** può essere usato anche con altri elementi qualora si renda necessario segnalare l'aspetto grafico del testo in essi contenuti. In tutti questi casi i valori da associare sono i seguenti:

italic per il corsivo

bold per il grassetto

sc per il maiuscoletto

enlarged per la doppia spaziatura tra i caratteri.

Apici e pedici

Qualora nel testo siano presenti caratteri o sequenze di caratteri in apice o pedice essi vanno codificati mediante l'elemento `<hi>` dotato di un attributo `type` con i seguenti valori:

- `apice`
- `pedice`

Ad esempio:

```
l<hi rend='apice'>am</hi>
```

Interventi editoriali

Se la fonte cartacea adottata per la digitalizzazione è una edizione critica o una edizione diplomatica che presenta nel testo indicazioni di congetture, lacune, correzioni, o altri interventi editoriali tali fenomeni vanno codificati usando gli elementi di base TEI per gli interventi editoriali. Naturalmente l'interpretazione della simbologia adottata in ciascuna edizione va effettuata leggendo con attenzione la nota editoriale del testo originale.

Per la codifica di omissioni o aggiunte nel testo sono disponibili gli elementi ``, `<add>` e `<gap>`.

L'elemento `` contiene una porzione di testo che appare cancellata sulla fonte per responsabilità dell'autore o di un copista. I seguenti attributi permettono di specificare la natura dell'omissione:

resp: indica l'istanza responsabile della individuazione dell'omissione. Se la responsabilità (come quasi sempre nel caso di fonti secondarie) del curatore il suo valore sarà `ed`.

rend: denota la forma materiale della cancellatura.

L'elemento `<add>` contiene invece porzioni di testo aggiunte dall'autore o da un copista. Oltre agli attributi del precedente esso è dotato dell'attributo `place`, attraverso il quale è possibile indicare la posizione fisica dell'aggiunta nella fonte originale del testo codificato: valori possibili di questo attributo possono essere stringhe come "margine sinistro", "infralineare" etc.

Il testo aggiunto o omissso codificato tramite le precedenti marche deve essere almeno parzialmente leggibile. Se si è in presenza di una vera e propria lacuna materiale (abrasione, cancellatura illeggibile, o se si tratta di fonti orali registrate, passi non ascoltabili) occorre usare piuttosto l'elemento vuoto `<gap>`. I seguenti attributi possono essere usati per determinare ulteriormente, ove possibile, la lacuna:

desc: fornisce una descrizione verbale della lacuna;

reason: indica le ragioni dell'omissione, tramite stringhe come "illeggibile", "cancellato", etc.;

Ad esempio il seguente dattiloscritto autografo degli *Appunti per "L'uomo e' forte"* di Corrado Alvaro alla fine del primo foglio presenta una cancellatura illeggibile,

```
ma come si prestava il fascismo. Non suscitano all'estero l'indigna-  
desc="sequenza di caratteri cancellata tramite sovrabbattitura di lettere x"  
reason="illeggibile" resp="ed">
```

```
<PB n='2' />
```

```
zione suscitata da fascismo, non fanno leva su un sentimento nazio-  
<lb/>
```

```
nale per la difesa e l'accrescimento d'una civilt&agrave; e d'una cultura,<lb/>
```

Si noti che l'elemento `<gap>` può in questo caso essere a sua volta contenuto da un elemento ``, poiché siamo in presenza di una soppressione da parte dell'autore.

L'elemento `<corr>` contiene una sequenza testuale che è stata corretta o emendata rispetto alla fonte originale dal curatore, in corrispondenza di un probabile errore. Si noti che `<corr>` può anche

essere usato per trascrivere nella copia elettronica di un testo critico a stampa gli emendamenti del curatore, o per registrare le correzioni introdotte da un copista in un manoscritto, quando questa sia individuabile. L'attributo **sic** permette di specificare (quando possibile) il testo originale errato o inesatto, mentre l'attributo **resp** svolge la stessa funzione vista negli elementi precedenti.

Liste, elenchi e glossari

In alcuni casi porzioni di testo, contenute nel testo vero e proprio o nelle sezioni peritestuali, si presentano nella forma di elenchi dotati di una struttura interna. A titolo di esempio si possono citare sommari, glossari, indici, liste lessicali, etc. Per la codifica di questi porzioni di testo sono previsti i seguenti elementi.

<list>: contiene una porzione di testo organizzata come una lista di items. Questo elemento non è minimizzabile, e può essere annidato entro se stesso. L'attributo **type** può essere usato per caratterizzare la lista in qualche modo rilevante. Il valore deve essere una sequenza di caratteri del tipo "glossario", "indice", etc.

<item>: contiene un singolo componente di una lista. Questo elemento è minimizzabile. Anche qui si può utilizzare l'attributo **type** che caratterizza il tipo di item.

<label>: contiene una etichetta che identifica ciascun componente di una lista

L'elemento **<list>** può contenere al suo inizio un elemento **<head>** che codifica il titolo della lista o dell'elenco. Attraverso questi elementi è possibile codificare un glossario: ogni entrata viene codificata dall'elemento **<label>**, mentre la relativa spiegazione può essere delimitata dall'elemento **<item>**.