

Lou Burnard
C.M. Spergerg-McQueen
**Il manuale TEI Lite.
Introduzione
alla codifica elettronica
dei testi letterari**

a cura di Fabio Ciotti, Milano,
Edizioni Sylvestre Bonnard, 2005
p. 268, ISBN 88-896009-00-1

L'efficacia e la chiarezza espositiva del manuale TEI Lite – la cui versione elettronica è disponibile liberamente all'indirizzo <http://www.tei-c.org/Lite/teiu5_it.xml>¹ – hanno indotto Fabio Ciotti a proporlo all'editore Sylvestre Bonnard per la pubblicazione a stampa. “Questo nella convinzione che il pubblico degli studiosi di scienze umane, degli archivisti, dei bibliotecari, ma anche quello dei professionisti nel settore editoriale, cui questo libro si rivolge, debba avere una conoscenza almeno introduttiva di uno strumento che si pone come inevitabile punto di riferimento per la digitalizzazione di testi e di documenti in ambito umanistico”, come Ciotti stesso dichiara a pagina 42 del saggio *La codifica del testo, XML e la Text Encoding Initiative* posto in apertura del volume di cui ci stiamo occupando.

Effettivamente, lo schema di codifica della TEI, basato sulla sintassi XML, e il mo-

numentale manuale che lo descrive, le *Guidelines for electronic Text Encoding and Interchange*,² pubblicata per la prima volta nel 2002 dal TEI Consortium³ a cura di Lou Burnard e C.M. Sperger-McQueen, sono dei veri e propri punti fermi per chiunque in ambito umanistico intenda costruire archivi e banche dati testuali con finalità e valore scientifici. I documenti originali prodotti dal TEI Consortium sono stati così tradotti da Ciotti e pubblicati dall'editore Sylvestre Bonnard con autorizzazione e licenza esclusiva del TEI Consortium. Il volume si apre con *Una semplice introduzione a XML*, che traduce *A gentle introduction to XML*, capitolo 2 delle *Guidelines for electronic Text Encoding and Interchange*, che in virtù della sua chiarezza nell'introdurre la sintassi di XML vanta già un'ampia diffusione anche come documento autonomo. Segue *TEI Lite: introduzione alla codifica dei testi*, traduzione italiana dell'agile manuale *TEI Lite: an introduction to Text Encoding for Interchange*,⁴ dedicato a un sottoinsieme dello schema di codifica TEI, il TEI Lite appunto, che è stato sviluppato appositamente per promuovere l'uso di questo linguaggio da parte di chi è interessato alla sperimentazione di tecnologie informatiche per le discipline umanistiche ma non è ancora in possesso delle competenze necessarie per affrontare lo schema nella sua totalità. Infine *La TEI Lite dalla P4 alla P5: continuità e cambiamento* traduce uno scritto originale di Lou Burnard (*TEI Lite from P4 to P5: continuity and change*) realizzato appositamente per questo volume. Nella letteratura tecnica si

definiscono linguaggi di codifica testuale ovvero linguaggi di marcatura (*markup language*)⁵ quei formalismi che consentono la rappresentazione formale di caratteristiche grafiche, strutturali o semantiche di un testo o di segmenti di testo. Si tratta di veri e propri linguaggi informatici, regolati da un insieme di caratteristiche testuali, da un insieme di identificatori simbolici (che separano le istruzioni di un linguaggio di markup dal testo verbale circostante e vengono denominati *tag*, marcatori), da una correlazione di identificatori e caratteristiche testuali, nonché da una sintassi che regola il modo in cui gli identificatori devono essere inseriti nel testo, sotto forma di marcatori, per segnalare l'occorrenza di una data caratteristica. Sia SGML (Standard Generalized Markup Language) che XML (Extensible Markup Language) sono metalinguaggi attraverso i quali è possibile definire formalmente un linguaggio di markup, cioè prescrivere precise regole sintattiche per definire un insieme di marcatori e le loro reciproche relazioni (senza dire nulla invece per quanto riguarda la loro tipologia, quantità e nomenclatura). La struttura di una classe di documenti XML può essere specificata esplicitamente mediante la definizione di una vera e propria grammatica formale, denominata Document Type Definition (DTD). Una volta definito formalmente, un linguaggio di markup XML può essere utilizzato per creare singoli documenti, che ne dovranno rispettare i vincoli grammaticali, oltre a conformarsi alle norme generali di sintassi XML. Tra le applicazioni del linguaggio XML più rilevanti e

complesse c'è senza dubbio lo schema di codifica cui questo volume è dedicato: la Text Encoding Initiative (TEI), un progetto che nasce negli anni Ottanta nell'ambiente degli studiosi di scienze del testo e in quello degli archivisti e dei bibliotecari, tutti ugualmente interessati a risolvere i problemi di interscambiabilità delle risorse digitali e di definire uno standard per la rappresentazione dei testi su supporto digitale. Nel 1988 il progetto TEI fu di fatto avviato e a partire dal 1989 la TEI ha sviluppato diverse versioni del suo schema di codifica, fino alla più recente, la TEIP5, introdotta lo scorso anno. Lo schema è stato formalizzato secondo una DTD che costituisce la grammatica descrittiva della classe di documenti appartenenti al suo dominio di riferimento, quello della ricerca umanistica. L'insieme delle caratteristiche testuali che costituiscono l'ontologia della TEI si articola su tre livelli: il primo è costituito da un insieme di caratteristiche testuali considerate come universalmente valide per tutti i documenti del dominio; il secondo è quello delle strutture testuali proprie di un limitato numero di sottoclassi individuate in base a macro-generi (testi in prosa, testi in versi, testi drammatici, vocabolari e dizionari, trascrizioni di fonti orali); il terzo è quello delle proprietà e delle caratteristiche testuali prodotte da singole prospettive analitiche (trascrizione diplomatica di fonti primarie, descrizioni di apparati di varianti, descrizione di strutture morfosintattiche, rappresentazione di strutture interpretative profonde, rappresentazione di strutture ipertestuali). Il risultato di questa impo-

stazione è una DTD che conta complessivamente oltre cinquecento elementi e cento attributi XML, la cui documentazione stampata occupa oltre duemila pagine. È così che per aiutare chi desidera avvicinarsi a questo complesso linguaggio di codifica i curatori della TEI hanno deciso di creare una versione "ristretta" della DTD e di redigere un manuale introduttivo che la documentasse: questa versione della DTD è stata battezzata TEI Lite e il suo manuale introduttivo è quello che ci viene oggi qui offerto in traduzione italiana. Il volume è corredato di due appendici. La prima, *Risorse informative e strumenti di elaborazione XML: una panoramica*, scritta da Fabio Ciotti, fornisce una rassegna delle risorse e degli strumenti software liberamente disponibili che possono essere usati per creare ed elaborare documenti in formato XML. Va precisato che, siccome il proliferare di strumenti e di programmi dedicati a XML rende oggi impossibile fornire una panoramica esaustiva di tutte le risorse disponibili, l'autore ha giustamente scelto di individuare quelle soluzioni che sembrano più adeguate in termini di efficacia, facilità di utilizzo e accessibilità, dividendole per tipologie applicative (risorse informative e documentazione online; strumenti per creare documenti XML; strumenti per visualizzare e pubblicare documenti XML; strumenti per interrogare e analizzare documenti XML) e lasciando ai lettori il compito di orientarsi autonomamente nel mondo della codifica XML. La seconda appendice, *Sull'utilizzo della codifica TEI in filologia*, di Simone Albonico, descrive al-

cune delle possibilità che lo schema TEI mette a disposizione per rappresentare fenomeni filologici e di critica testuale. In particolare, il saggio di Albonico intende mostrare attraverso alcuni esempi come il sistema di codifica TEI sia applicabile anche a situazioni complesse, tipiche nella tradizione degli studi testuali, in particolare italiana. L'autore si pone l'obiettivo di offrire una "piccola verifica, su un piano più tecnico, della validità e delle possibilità di applicazione di un sistema generale di codifica dei testi che, se ambisce a proporsi come soluzione standard, deve anche mostrare la sua duttilità ed estensibilità a situazioni non si dica estreme ma certo fortemente caratterizzate". Chiudono il libro la *Bibliografia* e l'*Indice generale*.

Roberta Cesana

Dipartimento di scienze
della storia
e della documentazione storica
Bibliografia e biblioteconomia
Università degli studi di Milano
roberta.cesana@guest.unimi.it

Text Encoding for Interchange,
TEI Consortium, 2002.

⁵ L'espressione "markup" deriva proprio dall'analogia che questi sistemi di codifica hanno con la simbologia e le annotazioni inserite da autori, curatori editoriali e correttori nei manoscritti e nelle bozze di stampa di un testo al fine di indicare correzioni e trattamenti editoriali.

Note

¹ <http://www.tei-c.org/Lite/teiu5_it.html>, per la versione in formato HTML.

² *Guidelines for electronic Text Encoding and Interchange*, edited by C.M. Sperberg-McQueen and L. Burnard, *XML-compatible edition* prepared by S. Bauman, L. Burnard, S. De Rose, and S. Rahtz, Oxford – Providence – Charlottesville – Bergen, TEI Consortium, 2002, p. 12-34.

³ Il TEI Consortium è un'organizzazione internazionale non-profit che gestisce e promuove la Text Encoding Initiative e detiene il copyright su tutte le sue specifiche tecniche e la relativa documentazione. Ulteriori informazioni sono disponibili all'indirizzo web del consorzio: <<http://www.tei.c.org/>>.

⁴ L. BURNARD – C.M. SPERBERG-MCQUEEN, *TEI Lite: an introduction to*