

Programmi di scrittura elettronica come Microsoft Word **ci hanno abituato a confondere il testo (il contenuto) con la sua veste grafica (la forma)**. Si prenda in esame il caso di un articolo scientifico redatto in vista di una pubblicazione su una rivista di settore.

Per far comprendere al lettore che una certa frase rappresenta il titolo di un singolo capitolo dell'articolo, ci preoccuperemo di metterla in rilievo assegnandole caratteri dalle dimensioni maggiori rispetto al corpo del testo (ma minori del titolo dell'articolo) e magari richiederemo al programma di centrarla nella pagina. L'impaginatore dovrà scorrere tutto l'articolo e individuare man mano i vari sottotitoli rimuovendo le formattazioni dell'autore (che però sono state funzionali all'individuazione dei sottotitoli in quanto tali) e assegnando le caratteristiche di formattazione proprie della rivista. In articoli complessi e con sottotitoli su più livelli questa operazione, oltre ad essere impegnativa e faticosa, comporta anche una certa possibilità di sviste ed errori.

Un autore più smaliziato avrebbe probabilmente fatto ricorso agli stili, assegnando uno stile "sottotitolo" a ciascun sottotitolo. Non essendo però diffuso alcuno standard sui nomi da assegnare agli stili in relazione alla loro funzione, l'impaginatore dovrà comunque riassegnare lo stile "sottotitolo" dell'autore allo stile funzionalmente equivalente in uso nella rivista che, trovandoci magari in un contesto internazionale, si chiama, poniamo, "subtitle". Gli basterà però farlo una sola volta per tutti i sottotitoli. In HTML (Hyper Text Markup Language) invece sarebbe stato sufficiente marcare il testo in questione con l'etichetta H2 (cioè *Heading*, "intestazione", di secondo livello).

Università "La Sapienza", Roma

L'Extensible Markup Language XML ci dà la possibilità di spingerci oltre. Normalmente, per segnalare al lettore che stiamo citando le parole di un altro autore, ci limitiamo quasi inconsciamente a farle precedere e seguire da una coppia di virgolette. **Con XML la nostra preoccupazione sarà invece quella di marcare tali parole con un'apposita etichetta;** poi potremo decidere (e altrettanto facilmente cambiare idea) che tutte le frasi così contrassegnate debbano apparire tra virgolette doppie, o caporali, o semplicemente in corsivo.

All'interno di un articolo possono però comparire anche parole o frasi che hanno una determinata funzione o significato anche se non sempre ciò corrisponde ad una particolare formattazione grafica. Ad esempio, in un articolo che si occupa di etimologie ugro-finniche potrebbe essere interessante marcare tutte le parole finlandesi e fare altrettanto con quelle ungheresi, in modo da poter creare poi rapidamente un indice analitico suddiviso in sezioni corrispondenti alle lingue citate. Similmente in qualsiasi tipo di articolo sarebbe sicuramente utile contrassegnare i nomi propri degli autori citati, oppure indicizzare le citazioni dei passi delle opere.

Prendiamo infine in considerazione la bibliografia, anche come esemplificazione di **XML applicato ad un insieme di dati indipendenti da un flusso di testo** (in contrasto quindi con gli esempi fin qui considerati). Ben prima che il testo dell'articolo abbia raggiunto una forma redazionale stabile, la bibliografia è innanzitutto uno strumento di lavoro ad uso dell'autore. Spesso sarebbe comodo poterla riordinare al volo secondo un criterio cronologico, poi tornare a quello alfabetico

Dottorati in Filologia greca e latina Linguistica storica Studi sul Vicino Oriente antico

per gli Studi Umanistici

secondo il cognome dell'autore; oppure segnalare le opere che abbiamo già consultato direttamente, o annotare la loro collocazione nella nostra biblioteca. Con XML possiamo creare una struttura in cui inserire dati quali i singoli autori, il loro ruolo rispetto a categorie quali traduzione e cura, il titolo, la serie, la rivista, il numero di volumi, l'anno di pubblicazione e qualsiasi altro dato bibliografico riterremo opportuno. In un secondo momento, attraverso una determinata tecnologia XML (detta eXtensible Stylesheet Language Transformations, XSLT), potremo fornire all'applicazione le regole con cui vogliamo che i dati vengano "trasformati" in modo funzionale ad una loro determinata presentazione. Ad esempio, sarà possibile riordinare cronologicamente la bibliografia in un attimo, visualizzare o meno le annotazioni personali, oppure elencare solo le opere pubblicate dopo un certo anno su una certa rivista, o mettere tutti gli anni tra parentesi e i titoli in corsivo, eliminando fra l'altro possibilità di sviste e non lasciando dubbi se un certo numero rappresenti il volume o la pagina.

Si immagini infine il caso di un testo teatrale dove siano marcate le singole parti dei diversi attori e distinte le indicazioni di scena.

XML è un linguaggio immediato e flessibile; **ciò che a volte può risultare difficile è la riflessione, che XML ci costringe a fare, sulla natura delle informazioni che vogliamo catalogare, strutturare, memorizzare o anche solo semplicemente scrivere e annotare.**

XML

1	Introduzione all'informatica umanistica ♦ Definizione ♦ Problematiche e finalità	lunedì	14.30-16
2	Introduzione al linguaggio HTML ♦ Principi della marcatura testuale ♦ Attività: pubblicazione di una bibliografia in una pagina web	lunedì	16.30-18
3	La codifica UNICODE in Microsoft Word, HTML e XML Segni diacritici e scritture in caratteri non latini	martedì	16.30-18
4	Introduzione al metalinguaggio XML ♦ Come codificare dati con XML ♦ Attività: codifica di un catalogo bibliografico	mercoledì	14.30-16
5	La rappresentazione di dati XML tramite XSL-T ♦ Selezione e presentazione di dati XML ♦ Attività: consultazione on-line del catalogo bibliografico	mercoledì	16.30-18
6	Alcune applicazioni XML per la codifica di testi ♦ Text Encoding Initiative (TEI) ♦ XML System for the Textual and Archaeological Research (XSTAR) ♦ Open Text	giovedì	16.30-18
7	Cuneiform Text Markup Language Il CTML come applicazione XML. Problematiche poste dai testi cuneiformi nella loro pubblicazione elettronica. La soluzione proposta da CTML.	venerdì	14.30-16
8	Esempi di utilizzo di CTML. Strumenti CTML: l'analizzatore lessicale. Possibili estensioni di CTML: i sistemi alfabetici.	venerdì	16.30-18

Università "La Sapienza", Roma

Dottorati in Filologia greca e latina
Linguistica storica e storia linguistica italiana
Studi filologici e letterari sul Vicino Oriente antico e l'Iran pre-islamico

L'eXtensible Markup Language

per gli Studi Umanistici

I7-2I maggio 2004

Dipartimento di storia moderna e contemporanea
laboratorio di informatica
piano seminterrato

Gian Paolo **Renello**,
Università di Salerno
Gian Pietro **Basello**,
Università "L'Orientale", Napoli
Stefano **Buscherini**,
Università "Alma Mater", Ravenna

Prof. Carlo G. Cereti
Coordinatore del corso