

XML - TEI per la codifica dei testi

Esercitazioni con Emacs

21 dicembre 2007

Per cominciare

- Breve riassunto dei concetti di base
- Strumenti: blocco note o Emacs
Attenzione! Il programma usato è, in linea di principio, indifferente... conta il risultato

TEI Emacs

- Emacs è un programma di scrittura per utenti avanzati
- È relativamente complesso da imparare
- La versione che useremo (se ne trovano altre versioni sul sito TEI: www.tei-c.org) è predisposta per il lavoro con XML-TEI:
 - Contiene le DTD
 - Include un parser (programma per verificare la validità del file)
 - Inserendo `</` chiude automaticamente il tag aperto

Importante!

- Molti comandi *non* funzionano secondo lo standard Windows
- Per esempio, non esiste il “nuovo file”: si crea un nuovo file usando il comando “apri” e scrivendo il nome di un file ancora non esistente
- Nel dubbio, usate i menu in alto

Requisiti per la prova finale

- Prendere il file che sarà pubblicato sulle pagine del corso
- Inserire al suo interno una poesia italiana, a scelta, inclusi i metadati richiesti
- Controllare come minimo che il file sia **ben formato** (va bene anche Internet Explorer)
- Spedirmelo per verifica: il file dovrà essere anche **valido** (è molto difficile renderlo non valido...)
- Non è necessario controllare che il file sia valido, ma se lo fate, tanto meglio

Poesia?

- Nel modello che vedrete sarà richiesto l'inserimento di un titolo
- Al di là di questo, il mattone fondamentale dei testi poetici nel sistema TEI è il verso, incluso all'interno di un elemento `<l>` (= *line*):
`<l>Nel mezzo del cammin di nostra vita</l>`
- Le strofe possono essere incluse in `<lg>` (= *line group*), elemento che può contenere più `<l>`

Lettere accentate, apostrofi...

- Subito un problema pratico!
- I caratteri accentati non sono supportati da tutti i sistemi informatici
- Ragioni storiche: gli standard per lo scambio dati sono stati inventati per la lingua inglese, e in inglese non ci sono caratteri accentati
- I caratteri accentati sono inseriti tramite espedienti di diverso livello, ormai *quasi* universali
- E all'italiano è andata ancora bene...

• ㄹㄹ ㄹㅇ ㅁㅁ ㅂㅇ ㅅㅇ ㅅㅁ ㅅㅂ ㅅㅅ ㅅㅇ ㅅㅇ !

Ascii di base

- ◆ Ascii ristretto, o a 7 bit
- ◆ Standard Iso 646
- ◆ Comprende 95 caratteri stampabili (lettere maiuscole e minuscole) e 33 caratteri di controllo
- ◆ Come Irv (International Reference Version) è tuttora l'unico assieme di caratteri universalmente condiviso
- ◆ Non contiene le lettere accentate!

Insiemi di caratteri a 8 bit

- ◆ Comprendono 256 caratteri
- ◆ Ne esistono diverse versioni
- ◆ Il più usato è l'Iso 8859-1 (“Latin1”), che comprende i caratteri più usati delle principali lingue occidentali
- ◆ Contengono le lettere accentate!
- ◆ In questi casi non è garantito l'interscambio

Unicode

- ◆ Sincronizzato (ma non coincidente) con lo standard Iso/Iec 10646
- ◆ <http://www.unicode.org>
- ◆ Per tutti (più o meno) gli alfabeti e per molte scritture non alfabetiche
- ◆ A ogni carattere viene assegnato un codice univoco
- ◆ Tutti i sistemi che supportano Unicode sono in grado di gestire questi codici

Nella pratica però...

- Nell'ortografia italiana si usano solo due accenti: grave (`) e acuto (/) - in poesia *potreste* trovare il circonflesso (*lambîr*)
- Gli accenti possono essere posti solo su vocali
- TEI lite consente l'uso di "entità" per indicare questi accenti:
 - &(lettera)grave; (accento grave)
 - &(lettera)acute; (accento acuto)
- N.B. Anche l'apostrofo è un carattere estraneo all'Ascii di base!

Obbligatorio per il compito

- Inserire gli accenti con il sistema appena visto (in modo che vengano visualizzati)
- Fare attenzione a eventuali copia e incolla, etc.

E per finire: CSS

- Cascading Style Sheet
- Forniscono indicazioni su come devono essere visualizzati i file
- Le indicazioni possono anche essere estremamente dettagliate
- Tutto si appoggia sulla marcatura XML