UNIVERSITÀ DI PISA

CODIFICA DI TESTI

ANGELO MARIO DEL GROSSO

Academic year	2019/20
Course	INFORMATICA UMANISTICA
Code	299LL
Credits	6

Modules	Area	Type	Hours	Teacher(s)
CODIFICA DI TESTI	L-FIL-LET/12	LEZIONI	42	ANGELO MARIO DEL GROSSO

Obiettivi di apprendimento

Conoscenze
Il corso si propone di fornire agli studenti la metodologia e gli strumenti necessari per effettuare la codifica scientificamente curata di testi, in particolar modo di testi letterari o comunque di interesse storico-culturale, adottando gli schemi XML sviluppati nell'ambito della Text Encoding Initiative (TEI). La parte iniziale del corso sarà dedicata a una riflessione teorica introduttiva delle problematiche di questa materia.

Modalità di verifica delle conoscenze
Al fine di verificare le conoscenze acquisite dagli studenti del corso saranno svolte, anche in itinere, degli esercizi di trascrizione e codifica di documenti testuali, in parte autonomamente, in parte come esercitazione sotto la guida del docente. Tali prove non riceveranno una valutazione effettiva, ma ogni studente riceverà opportune indicazioni e suggerimenti su eventuali errori e potenziali soluzioni. Lo svolgimento degli esercizi è obbligatorio per il sostenimento dell'esame, in quanto alcuni esercizi potrebbero essere ripresi e commentati in sede d'esame.

Capacità
Al termine del corso lo studente sarà in grado di analizzare il documento da codificare, di valutare il metodo di codifica più appropriato, di creare uno schema di codifica TEI e di usare gli strumenti software più idonei (editor XML, XSLT etc.) per la codifica, la trasformazione e la fruizione di un testo digitale.

Modalità di verifica delle capacità
Lo studente dovrà preparare e presentare un progetto di codifica relativo ad un testo **concordato**, a scelta dello studente o proposto dal docente, che mostri le capacità di codifica e di visualizzazione apprese durante il corso. Il progetto deve essere obbligatoriamente caricato sulla piattaforma github una settimana prima della sessione d'esame.

Comportamenti
Lo studente imparerà e/o svilupperà abilità nell'ambito delle attuali metodologie e delle migliori pratiche relative alla codifica di testi, con particolare attenzione alle linee guida sviluppate dalla Text Encoding Initiative (TEI).

Modalità di verifica dei comportamenti
Durante il corso, il docente valuterà le competenze acquisite dagli studenti sulle tecnologie XML e sulle pratiche nella codifica di testi con riferimento alle linee guida della Text Encoding Initiative (TEI).

Prerequisiti (conoscenze iniziali)
Non ci sono prerequisiti obbligatori, ma la conoscenza dei rudimenti di HTML/CSS e di altre tecnologie del Web è fondamentale per implementare quanto appreso durante il corso.

Indicazioni metodologiche
Il corso si terrà sotto forma di lezioni frontali con l'ausilio di slide e consultazione di specifici siti web. Il docente metterà a disposizione degli studenti le slide e gli esercizi che saranno svolti a lezione caricando i materiali sullo spazio github del docente. La comunicazione con gli studenti, in particolare per quanto riguarda gli esercizi , avverrà prevalentemente a lezione e via email.

Programma (contenuti dell'insegnamento)

Gli argomenti trattati saranno i seguenti:

- introduzione teorica alla codifica di testi
- i linguaggi di markup e fondamenti del linguaggio XML
- concetti di base per la creazione di schemi di codifica con DTD e XML Schema
- le norme TEI (Text Encoding Initiative) per la codifica di testi di vario tipo e genere
- introduzione al modulo TEI dictionary, e ai principali elementi per edizioni in facsimile ed edizioni digitali diplomatiche ed interpretative
- definizione di schemi di codifica TEI personalizzati
- introduzione ai fogli di stile XSLT e CSS
- introduzione al Document Object Model per il trattamento di documenti TEI-XML

Bibliografia e materiale didattico

Oltre alle diapositive delle lezioni, messe a disposizione dal docente sulla piattaforma github e Moodle, i seguenti testi:
TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [3.6.0] [Agosto 2019]. TEI Consortium. http://www.tei-c.org/Guidelines/P5/
Sul sito TEI: http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html, disponibili sia in formato HTML per la consultazione online, sia in formato PDF e ebook (ePub/Mobi).
Libro: What is the Text Encoding Initiative? How to add intelligent markup to digital resources scritto da Lou BURNARD. Nouva edizione [online]. Marseille: OpenEdition Press, 2014. Disponibile su Internet: <http://books.openedition.org/oep/426>. ISBN: 9782821834606. DOI: 10.4000/books.oep.426.
Libro: Elena Pierazzo. Digital Scholarly Editing: Theories, Models and Methods. 2014. ?hal-01182162?
Un capitolo a scelta, in base al tipo di progetto, di *Electronic Textual Editing*, eds. Lou Burnard, Katherine O'Brien O'Keeffe, John Unsworth. Modern Language Association of America, 2006. Disponibile online (versione preview): http://www-old.tei-c.org/About/Archive_new/ETE/Preview/.
Importante: le diapositive delle lezioni relative agli elementi TEI XML costituiscono soprattutto una guida in merito ai relativi capitoli delle *Guidelines*, dove la materia trattata deve essere studiata e approfondita.
Altri testi di riferimento e approfondimento verranno comunicati a lezione.

Indicazioni per non frequentanti

Per i non frequentanti si richiede di contattare il docente, anche via email (angelo.delgrosso@ilc.cnr.it), per concordare il programma, basato comunque sul normale programma d'esame salvo interessi particolari dello studente, e il progetto di esame.

Modalità d'esame

Esame orale su principi e metodi della codifica di testi, sarà inoltre discusso il progetto di codifica concordato con lo studente alla fine del corso.
Il progetto d'esame deve obbligatoriamente essere caricato sulla piattaforma github una settimana prima della data d'esame.
Per l'anno accademico in corso il docente ha messo a disposizione due corpora dai quali scegliere il materiale per il progetto d'esame.
Il primo corpus è un insieme di cartoline storiche le cui scansioni sono disponibili al seguente link:
< https://github.com/angelodel80/corsoCodifica/tree/master/progetto-esame/ProgettoCartoline/Scan_Cartoline_jpg >
Il secondo corpus è un insieme di lettere autografe di Vincenzo Bellini, le cui scansioni sono disponibili al seguente link:
< https://github.com/angelodel80/corsoCodifica/tree/master/CTaa19-20/progetto-esame/scans/LL1 >
I progetti d'esame possono essere preparati individualmente oppure in gruppo (da due -preferibile- massimo tre persone). Per le cartoline il progetto individuale prevede la scelta di due cartoline, mentre, in gruppo, il progetto prevede la codifica di tre cartoline. Per le lettere di Bellini la scelta dei materiali corrisponde al grado di difficoltà delle lettere indicato nel seguente file per ciascun caso (individuale, gruppo):
< https://github.com/angelodel80/corsoCodifica/blob/master/CTaa19-20/progetto-esame/appunti/grado_codifica_Lettere.docx >
Per entrambi i corpora la scelta dei materiali deve essere comunicata al docente prima di procedere con la codifica dei documenti, in più i documenti non devono essere presenti nella lista dei materiali già codificati in altri progetti, consultabile ai seguenti link:
(Cartoline)
< https://github.com/angelodel80/corsoCodifica/blob/master/progetto-esame/README.md >
(Lettere)
< https://github.com/angelodel80/corsoCodifica/blob/master/CTaa19-20/progetto-esame/README.md >

Stage e tirocini

Il docente è interessato a percorsi di stage e tirocini da compiersi all'interno dell'Istituto di Linguistica Computazionale "A. Zampolli" del CNR di Pisa in co-tutela con l'università.

Pagina web del corso

https://elearning.humnet.unipi.it/course/view.php?id=2970

Altri riferimenti web

Altri materiali del corso sono disponibili sulla pagina personale github del docente al seguente indirizzo:
https://github.com/angelodel80/corsoCodifica