

Dipartimento di Matematica, Università di Pisa

COSTRUZIONE DEL SOFTWARE PER L'EDIZIONE DIGITALE DEL *Liber  
Abbaci* CON APPLICAZIONE DI TECNICHE DI MACHINE LEARNING

**Relazione semestrale**

Michele Fioravera

2020, luglio-dicembre

Questo documento descrive l'attività svolta in seguito alla presa di servizio per la produzione di un'edizione critica digitale del *Liber Abbaci* di Leonardo Pisano. La ricerca segue le indicazioni del responsabile nel rispetto delle linee guida condivise con i partecipanti al progetto. L'indagine avviene collaborando prevalentemente con il gruppo di lavoro il cui obiettivo corrente è tradurre in TEI P5 l'intero oggetto di studio entro la conclusione del secondo semestre. Segue lo stato di avanzamento dei compiti assegnati contestualmente al conferimento dell'assegno.

**I risultati di manutenzione correttiva ed evolutiva degli strumenti informatici di supporto all'uso del linguaggio MauroTeX** riguardano principalmente il sistema di traduttori e sono documentati negli aggiornamenti al manuale utente e ai file di test.

Ho apportato modifiche ai desktop software che permettono di convertire la rappresentazione  $\text{\TeX}$  in formati adatti a svolgere efficientemente specifici scopi, seguendo le indicazioni del product owner e confrontandomi con il team di sviluppo. Le revisioni - comprese tra r221 e r347 - coinvolgono gli strumenti di analisi lessicale e di parsing: sono state introdotte nuove funzionalità di traduzione in  $\text{\LaTeX}$  e HTML per rispondere a necessità di stesura di bozze e di analisi stemmatica emergenti nel corso del progetto. Le operazioni di sviluppo sono iniziate durante le prime settimane seguendo le indicazioni del responsabile presso l'operatore della filiera culturale e creativa regionale, durante incontri occasionali svolti nel primo pomeriggio presso i locali dipartimentali. A partire da settembre la ricerca si è svolta di concerto con la cooperativa eLabor di Pisa secondo metodologia agile monitorata quotidianamente: il lavoro autonomo viene concordato tramite stand-up meeting coordinati dal product owner insieme ai componenti del team di sviluppo. La pratica è coadiuvata dall'adozione di piattaforme per l'organizzazione a distanza, la comunicazione sincrona e la gestione asincrona di bacheca SCRUM. La tabella 1 illustra un sommario dei lavori completati.

traduttore	modifica	inizio	fine
m2ledmac	eliminare paragrafi “vuoti”	Oct, 29	Nov, 14
	VV annidate nelle Cit		
	gestione “A+”		
m2hv	estrazione commenti		
m2ledmac	gestione VV	Sep, 28	Oct, 26
	gestione Cit	Sep, 25	
		stili note	
m2lv	gestione figura	Sep, 18	Sep, 30
m2ledmac	gestione paragrafi		
m2lv	aggiungere un ‘\par’ alla fine di protasis		
(tutti)	folium: verificare presenza del testimone ...		

Tabella 1: Traduttori coinvolti da modifiche nel quadrimestre settembre–dicembre.

La prima colonna elenca i traduttori coinvolti dalla manutenzione; la seconda riporta una sintesi della modifica. Le colonne “inizio” e “fine” riportano rispettivamente la prima data di aggregazione del mio profilo alla scheda e la data di eliminazione della scheda dalla bacheca - previa accettazione del team - da parte del product owner.

In seguito ai lavori sintetizzati dalla tabella 1 sono stati condotti i lavori dedicati al miglioramento dell’analisi stemmatica: l’ottenimento della lista delle concordanze interne ad un gruppo di testimoni contro quelle di un altro gruppo di testimoni, includendo possibilmente anche il testo base in uno dei due gruppi, è il tema della scheda che attende l’ultima fase di review. Sono stati successivamente condotti e sospesi i lavori riguardo la possibilità di produrre un sorgente MauroTeX di un dato testimone (rappresentati nel sistema di versionamento come nuovo programma di output: m2m).

La manutenzione del traduttore in linguaggio XML è prevista per il secondo semestre.

**Le possibilità di realizzazione di un ambiente integrato di sviluppo che permetta di gestire in modo semplice edizioni digitali del tipo di quella del Liber Abbaci** sono condivise sulla base delle esperienze relative ai sistemi EVT e Scriptorium.

Sono stato coinvolto principalmente nella risoluzione di problematiche relative alla sintassi del linguaggio di marcatura prescelto, con particolare riferimento alle pratiche di utilizzo degli strumenti di ‘editing’ da parte di utenti finali appartenenti all’ambito accademico e editoriale.

Ho presentato il tema della *preparazione di strumenti per un’uscita a stampa di qualità fruibile da singoli studiosi o case editrici* in sede di riunione plenaria (avvenuta dopo i primi quattro mesi seguenti la presa di servizio).

**Il contributo a strumenti di analisi avanzata dei testi con particolare riferimento a tecniche di 'machine learning' (ML) e di intelligenza artificiale** è limitato allo studio dello stato dell'arte relativo ai processi di 'natural language processing' (NLP).

Lo studio è stato introdotto per accogliere la richiesta di produrre per l'ultima settimana di agosto 2020

- un seminario che svolga una panoramica delle problematiche e tecniche NLP/ML più utilizzate;
- un workshop in cui condurre alcune persone alla realizzazione pratica di una delle tecniche presentate.

La presentazione dello studio è stata registrata come possibilità formativa e documento dell'attività svolta. Successive azioni di trasferimento tecnologico sono state caratterizzate da simili modalità condotte in parallelo con lo sviluppo: ho effettuato un 'tech pill' nel mese di dicembre a sintesi di studi preliminari circa strumenti utili al raggiungimento dell'obiettivo principale che coinvolgerà i lavori del secondo semestre; ho aderito all'iniziativa dei "corsi hands-on" promossa dal Laboratorio di Cultura Digitale, confrontando i miei studi con quanto proposto dal corso *Strumenti per l'estrazione di informazioni da testi*.

La ricerca bibliografica viene aggiornata periodicamente grazie a strumenti offerti dal sistema bibliotecario di ateneo. La partecipazione in qualità di uditore alle azioni divulgative proposte nel territorio e online integra gli studi.



Allegato n. 10  
Pag. n. 3  
Verbale del 18-02-2021



