REQUISITI

* Per il Primo parziale, quindi entro il 28 aprile, bisogna portare solo la documentazione (quindi senza nessun codice Java). Dato che la prof vuole che sia tutto in una repository git, organizzata in:

<nome\_progetto> : root del progetto

- src : questa directory conterrà il codice sorgente del progetto

- doc : questa directory conterrà i documenti di design del progetto (in format pdf)

- javadoc : questa directory conterrà il codice javadoc del progetto

Questo vuol dire che solo la cartella doc dovrà contenere roba, mentre le altre due dovranno essere ancora vuote.

* La documentazione dovrà essere salvata in formato PDF.

RIELABORAZIONE DELLA TRACCIA

Lo scopo di questo progetto è di consentire la consultazione dei manoscritti che devono essere digitalizzati e che costituiscono un patrimonio bibliografico antico per un totale di 60.000 carte (ms. sec. XV-XIX) contenenti memorie storiche della città dell’Aquila.

Sostanzialmente, bisogna creare un sistema che interagisca con un database di questi manoscritti. Ogni manoscritto segue questo ciclo di vita:

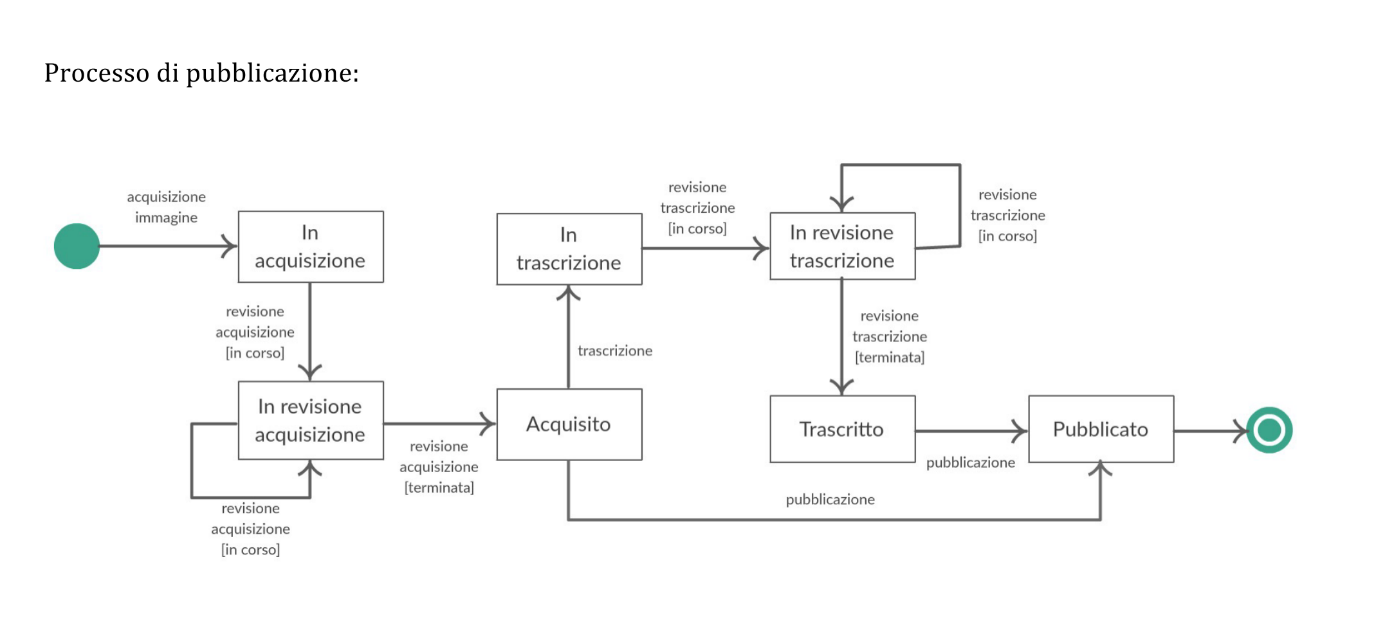
-Inizialmente, c’è solo l’opera, solamente registrata nel sistema, ma di cui non sono presenti né scansioni né trascrizioni. \*DUBBIO: magari un’opera che non ha né scan né trascrizioni dovrebbe essere omessa dalla ricerca, e quindi dal sistema. Questo vuol dire che un’opera viene caricata nel sistema solo quando sono presenti tutte le scansioni, ma questo potrebbe essere complicato. Una soluzione potrebbe essere mantenere l’opera nel sistema anche se non sono presenti scansioni, ma ometterla dai risultati della ricerca tramite metadati.\*

-Degli utenti con particolari permessi caricano le scansioni delle singole pagine dell’opera. I responsabili delle scansioni controllano le immagini inviate e decidono se vanno bene o meno.

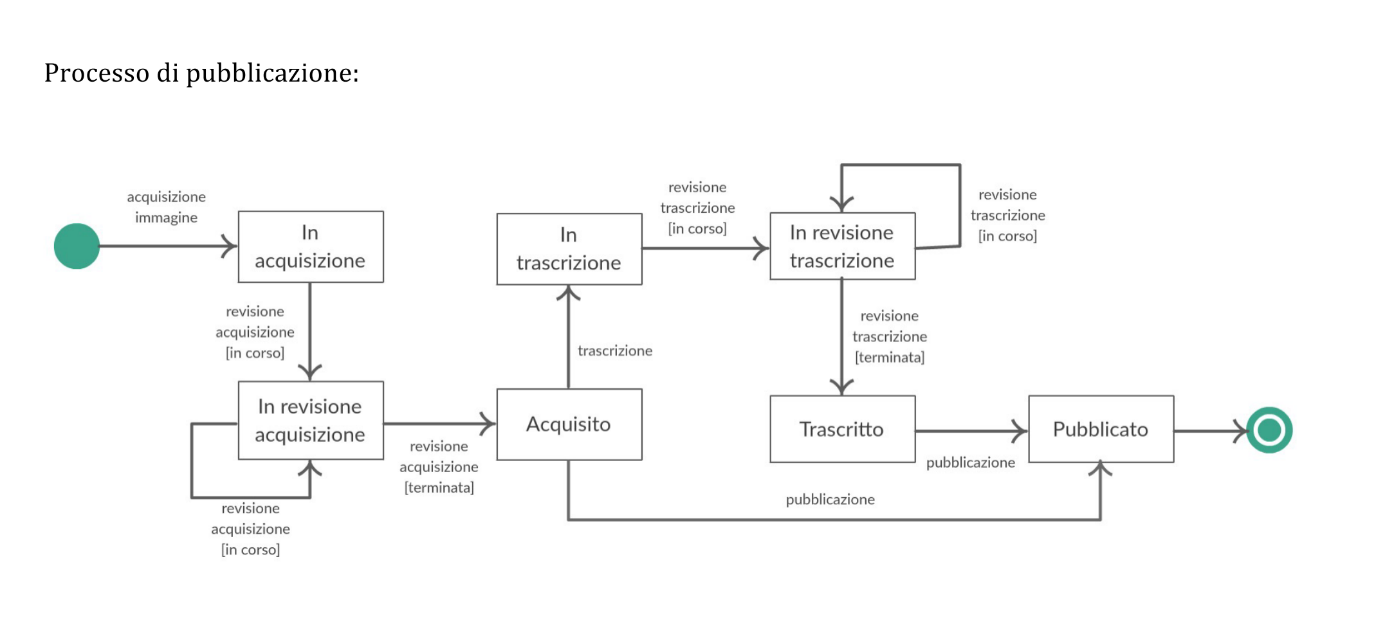
A questo punto l’opera è già consultabile dagli utenti registrati, e può essere cercata tramite metadati. L’utente non sarà in grado di vedere la trascrizione, perché non ancora presente.

-Una volta che l’opera è stata interamente scansionata, i trascrittori cominciano a trascrivere. E’ il manager del sistema ad assegnare una o più pagine a ciascun trascrittore. I responsabili delle trascrizioni (che sono diversi da quelli delle scansioni) controllano se la trascrizione va bene.

Una volta conclusa la trascrizione dell’intera opera, viene pubblicata e l’utente, cercando l’opera, troverà al fianco di ogni pagina scansionata la relativa trascrizione.

Ci sono alcune funzionalità minori che non ho menzionato, ma questo è in generale il ciclo di vita di ogni pagina, dal nulla alla scansione + trascrizione. Lo schema che ci ha dato la prof e che riassume questo processo è il seguente: 

Tuttavia, come la prof ci ha detto, questo schema è solo una bozza e non è da considerare completo. In effetti, possono essere fatte alcune modifiche:





La modifica 1 rimuove un ciclo non necessario: una volta che lo scan è in acquisizione, può essere accettato o rifiutato. Nel primo caso, si passa ad Acquisito. Nel secondo caso, si torna ad In acquisizione. Il ciclo è inutile e al suo posto viene aggiunta la freccia che gestisce il secondo caso.

La modifica 2 aggiunge una funzionalità importante anche se non espressamente scritta nel testo: un trascrittore può decidere che lo scan non è abbastanza leggibile e segnalarlo all’amministratore, che rimanderà dunque la pagina alla fase di Acquisizione scan. Dato che i documenti sono 60.000, ognuno con diverse centinaia di pagine, non è impensabile che il responsabile degli scan possa erroneamente accettarne uno che in realtà non va bene.

Il 3 non indica una modifica ma è una nota: al contrario di 1, qui il ciclo ha senso, perché una trascrizione può essere migliorata anche solo in alcune parti, al contrario della singola immagine, che invece può essere o accettata in toto o rifiutata in toto.

Ulteriori modifiche allo schema sono probabilmente necessarie.

Oltre al processo che viene fatto da ogni pagina di ogni manoscritto, il sistema presenta un sistema di Registrazione, Profilo e Login e di conseguenza diversi tipi di utenti con diversi privilegi (da qui si parte per fare gli use-cases). I tipi di utenti finora individuati sono:

* Utente registrato (utente base): può accedere, visualizzare il profilo, fare richiesta per diventare trascrittore, ricercare un manoscritto per metadati e visualizzarlo. Nota: dato che sono vecchi manoscritti, non è detto che tutti abbiano un titolo. Un utente deve poter cercare il manoscritto anche se non conosce il titolo, attraverso dei metadati che saranno attributi del manoscritto (esempio: anno, periodo, genere, ecc..)
* Uploader: l’uploader è colui che manda gli scan dei manoscritti in attesa di approvazione. Gli uploader devono essere in possesso dei manoscritti (per poterli scannerizzare) e di conseguenza si tratta di persone interne al progetto e non di utenti qualunque che vengono promossi. Di conseguenza, si potrebbe pensare agli uploader anche come utenti non registrati.
* Supervisore di scan: controllano le immagini mandate dall’uploader e decidono se vanno bene o meno. Non è detto che siano loro stessi uploader.
* Trascrittore: vengono incaricati dal manager di trascrivere una o più pagine. Possono inviare la loro trascrizione e aspettare il giudizio dei supervisori.
* Supervisore di trascrizione: controllano la correttezza delle trascrizioni. Sono diversi dai supervisori di scan; si tratta probabilmente di esperti storici e linguistici che controllano il contenuto.
* Manager: dà ai trascrittori le pagine da tradurre. Probabilmente è anche quello che gestisce il back-end del sistema, anche se la traccia non è molto chiara a riguardo.

Ci sono inoltre una particolare tipologia di utenti che può scaricare le opere. Io però questi non li metterei come utenti a sé stanti (Utenti Pro) ma farei invece in modo che il particolare privilegio di poter scaricare le opere possa esser dato a qualunque tipo di utente, indipendentemente dalla sua posizione e dal suo ruolo nel sistema.

---TO BE CONTINUED 🡪