

Progettazione

1 TEAM

1.1 COMPONENTI TEAM

- Barbanti Francesco MATR: 112374
- Canaris Andrea MATR: 121500
- Guerzoni Sharon MATR: 119901
- Ianniciello Andrea MATR: 126115
- Milana Riccardo MATR: 113180
- Morselli Daniel MATR: 121727

1.2 RUOLI

- Scrum Master: Barbanti Francesco
- Product owner: Canaris Andrea
- Developer team:
 - Front-end: Guerzoni Sharon, Milana Riccardo, Morselli Daniel
 - Back-end: Barbanti Francesco, Ianniciello Andrea
 - Tester: Barbanti Francesco, Ianniciello Andrea, Milana Riccardo
 - Full stack: Canaris Andrea
 - Documentazione: Canaris Andrea, Guerzoni Sharon, Morselli Daniel

1.3 METODOLOGIA DI SVILUPPO E STRUMENTI UTILIZZATI

La metodologia di sviluppo adottata è scrum.

Gli strumenti utilizzati per la creazione e condivisione del materiale sono:

- Trello (<https://trello.com/invite/b/eplBlQv8/31e3b53f2d05d10f3171e0ac0227f0f3/pds>): per la gestione delle user stories
- Pointingpoker.com: per la valutazione delle user stories
- GitLab
- Google Meet: per il lavoro di gruppo da remoto
- Google Drive: per la condivisione dei vari materiali
- WhatsApp: per organizzazione degli incontri

2 DESCRIZIONE DELLA SOLUZIONE PROPOSTA

2.1 CONTENUTI PRINCIPALI:

L'obiettivo principale del progetto "*Analizza il vocabolario*" è l'analisi ed il confronto della complessità di documenti di testo inseriti dai membri di una comunità online. L'implementazione del sito web sfrutterà Django, un framework web Python ad alto livello (software progettato per supportare lo sviluppo di siti web dinamici), open source (non protetto da copyright e liberamente modificabile dagli utenti) e free (ne viene concesso lo studio, l'utilizzo, la modifica e la redistribuzione).

I membri potranno usufruire dei servizi offerti dal sito web solo iscrivendosi alla comunità tramite un username univoco.

L'utente, dopo l'autenticazione delle sue credenziali, viene reindirizzato alla homepage del sito al cui interno è contenuta una breve descrizione di tutte le funzionalità di esso oltre che ad una breve guida per l'utilizzo degli strumenti.

Un membro registrato può inserire il testo in un'apposita text area del sito insieme ad altri elementi utili per la ricerca dei documenti quali: titolo del testo, l'autore, il tipo di documento (campi obbligatori) e opzionalmente una breve descrizione. In particolare, i tipi di testo individuati per la ricerca sono: tema scolastico, saggio, articolo di giornale, altro.

Dopo il caricamento del testo sul browser, attraverso la libreria Python NLTK, verranno rimossi i segni di interpunzione e tutti caratteri non attinenti.

Successivamente, con l'ausilio della stessa libreria, si procede ad una suddivisione delle singole parole (token), che saranno memorizzate all'interno del database con associata la loro frequenza nel testo.

Ad ogni documento T sarà assegnato un indice di complessità IC_T determinato dalla seguente formula:

$$IC_T = \frac{1}{10} * \left(\sum_{t_{unici} \in T} \frac{c_t}{f_t} \right) * dl_T$$

La complessità delle singole parole c_t corrisponde al numero di sillabe che possiedono, determinata sempre con l'ausilio di NLTK.

Come si può notare dalla formula, la complessità di un testo IC_T è determinata dalla somma della complessità delle singole parole c_t contenute in esso diviso per la loro frequenza f_t .

Successivamente, si moltiplica tale risultato per la densità lessicale del testo dl_T , ovvero:

$$dl_T = \frac{n^{\circ} \text{ parole uniche}_T}{n^{\circ} \text{ parole totali}_T}$$

Essendo una comunità, ogni utente iscritto può ricercare qualsiasi testo disponibile sul sito tramite le informazioni registrate durante la pubblicazione del documento (metadati) e visualizzarne le caratteristiche sopra specificate.

Inoltre, un utente registrato può confrontare due testi, visualizzando i rispettivi indici di complessità e frequenze delle parole.

Per memorizzare tutte le informazioni relative ai testi inseriti e agli utenti iscritti si ricorre ad un Database SQLite, la cui gestione avviene sempre attraverso il framework Django.

In aggiunta, dato un testo, si vuole dare la possibilità all'utente di apprendere i suoi termini in altre lingue, visualizzando la traduzione delle parole nella lingua selezionata. Quest'ultima operazione verrà realizzata mediante l'uso dell'API (insieme di procedure atte all'espletamento di un dato compito) di Google Translate.

Oltre a ciò, un utente registrato può richiedere l'eliminazione del proprio account, che ha come conseguenza l'eliminazione dei suoi testi inseriti dal database.

2.2 CONTENUTI AGGIUNTIVI

In questo paragrafo vengono trattati i contenuti opzionali e non richiesti esplicitamente dal cliente.

Si vuole dare la possibilità ad un utente della comunità di visualizzare i sinonimi delle parole presenti di un testo selezionato. In particolare, questa feature viene realizzata tramite la determinazione del contesto della parola all'interno della frase, seguito dal recupero dei suoi sinonimi utilizzando il thesaurus Wordnet attraverso la libreria NLTK.

Infine, per fornire una visione più estesa della complessità, si vogliono visualizzare: un indice di densità lessicale, numero di parole uniche e il numero di parole totali relativo al testo.

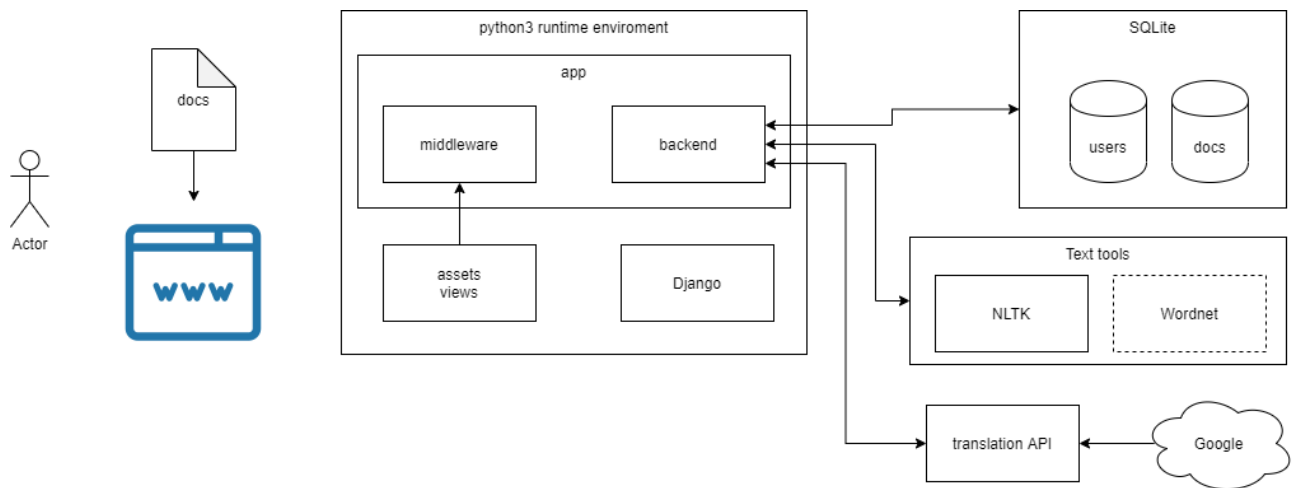
2.3 GLOSSARIO

- **API (Application Programming Interface):**
interfacce che sviluppatori e programmatori terzi possono utilizzare per espandere le funzionalità di programmi, applicazioni e piattaforme di vario genere.
- **Database:**
archivio di dati strutturato in modo da razionalizzare la gestione e l'aggiornamento delle informazioni e da permettere lo svolgimento di ricerche complesse.
- **DBMS (Database Management System):**
è un sistema software progettato per consentire la creazione, la manipolazione e l'interrogazione efficiente di database.
- **Django:**
Django è un web framework con licenza open source per lo sviluppo di applicazioni web, scritto in linguaggio Python
- **Framework:**
con framework si intende la struttura di un programma che viene utilizzata come base nello sviluppo software.
- **Librerie di linguaggio:**
una libreria è un insieme di funzioni o strutture dati predefinite e predisposte per essere importate in un programma software attraverso un opportuno collegamento.
- **Linguaggio ad alto livello:**
i linguaggi ad alto livello sono linguaggi informatici concepiti per facilitare il dialogo con l'operatore, facendo uso di istruzioni più complesse e potenti, e identificate da sigle tratte in genere dalla lingua parlata.

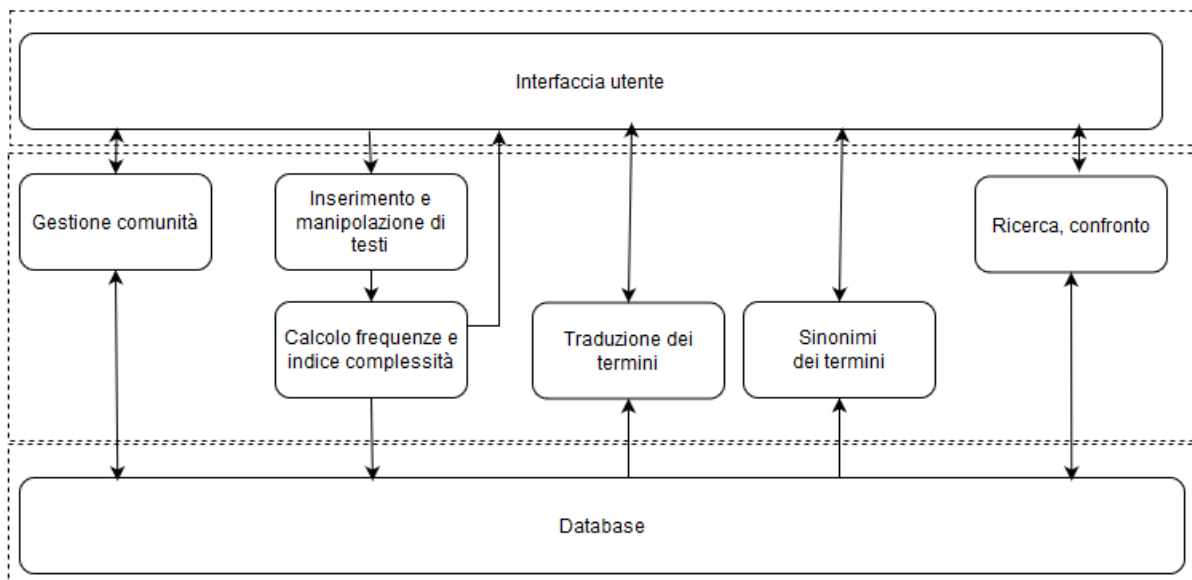
- **Metadati:**
i metadati sono le informazioni di cui bisogna dotare il documento informatico per poterlo correttamente formare, gestire e conservare nel tempo.
- **NLTK (Natural Language Toolkit):**
è un insieme di librerie e programmi scritti in linguaggio Python, per l'analisi simbolica e la statistica nel campo dell'elaborazione del linguaggio naturale.
- **SQLite:**
SQLite è un DBMS rilasciato con licenza libera.
- **Python:**
è un linguaggio di programmazione ad alto livello.
- **Software free:**
un software free (o libero) è un software distribuito sotto i termini di una licenza di software libero, che ne concede lo studio, l'utilizzo, la modifica e la redistribuzione grazie all'utilizzo di particolari licenze software.
- **Software open source:**
un software open source è reso tale per mezzo di una licenza attraverso cui i detentori dei diritti favoriscono la modifica, lo studio, l'utilizzo e la redistribuzione del codice sorgente.
- **Thesauro:**
un thesaurus (o tesoro) è il lessico dei termini relativi ad un ambito generale o specifico di conoscenze, collegati tra loro in una rete gerarchica e relazionale. Il termine ha anche il significato di vocabolario per opere lessicografiche, repertori scientifici, lessici storici.
- **Token:**
un token è un blocco di testo categorizzato, normalmente costituito da caratteri indivisibili chiamati lessemi.
- **WordNet:**
WordNet è un sistema di gestione di un dizionario semantico-lessicale basato sulle teorie psicolinguistiche della memoria lessicale umana. WordNet riconosce quattro categorie sintattiche: nomi, verbi, aggettivi ed avverbi, ognuna delle quali è organizzata in insiemi di sinonimi (synonym sets o synsets).

3 SCHEMI E DIAGRAMMI

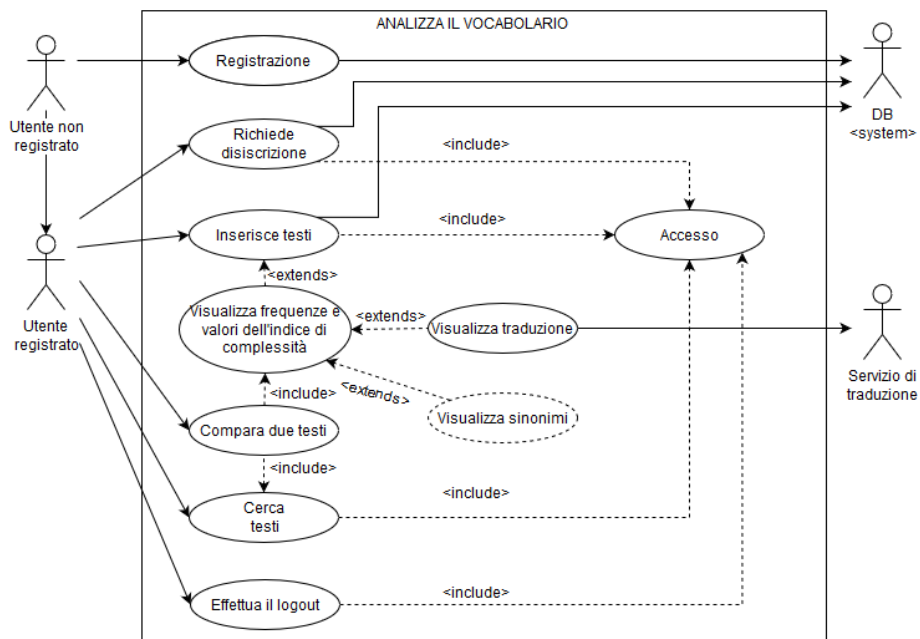
3.1 SCHEMA ARCHITETTURALE



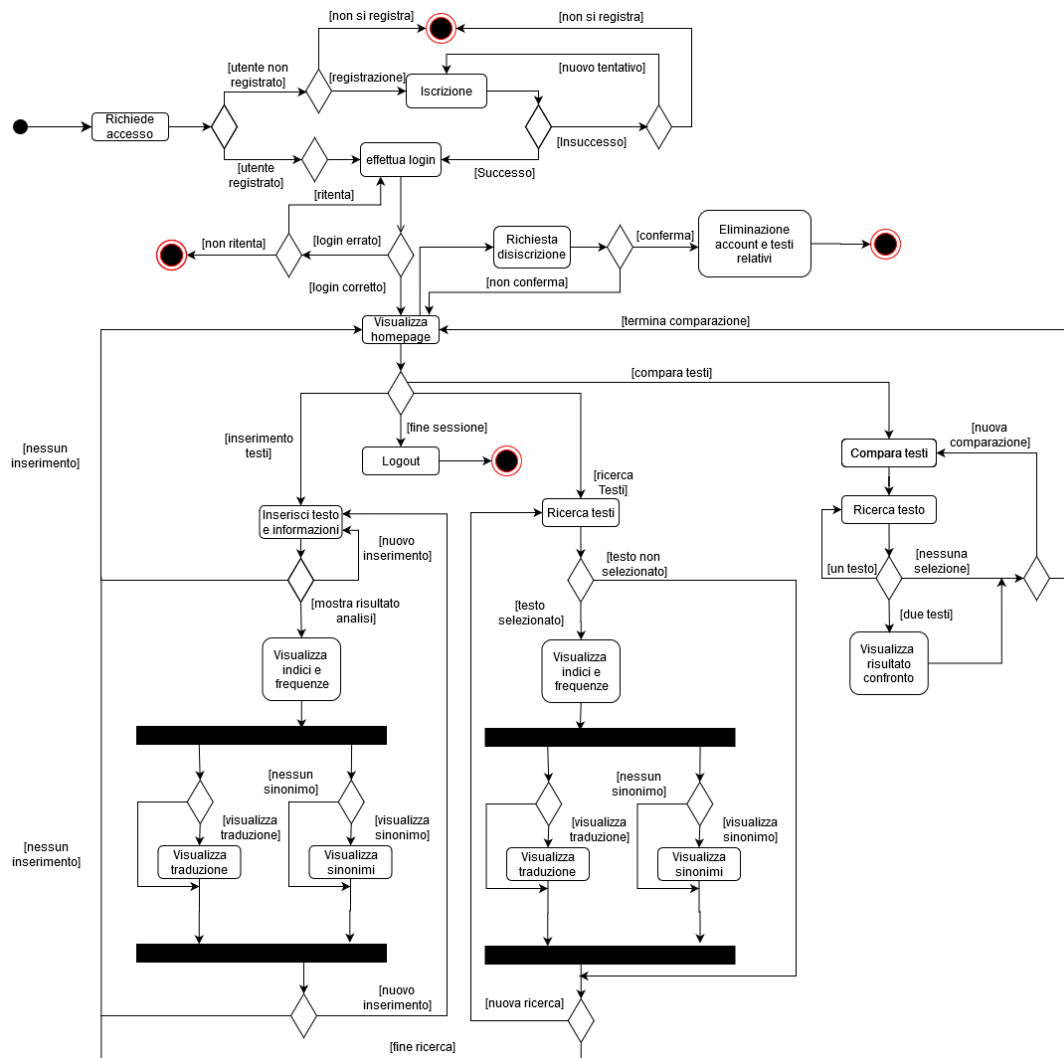
3.2 SCHEMA A BLOCCHI

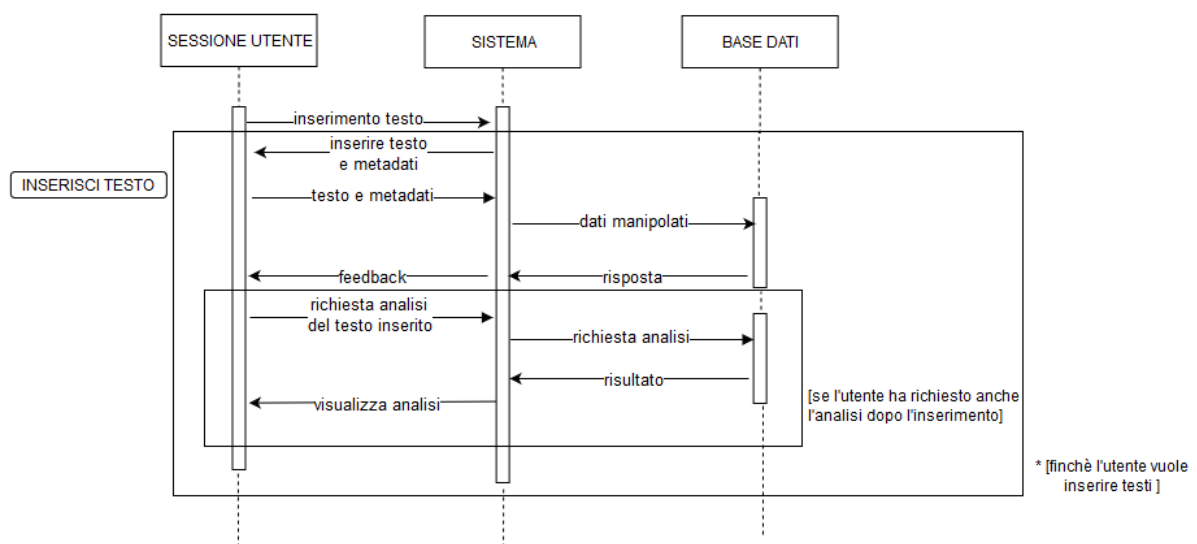


3.3 DIAGRAMMI DEI CASI D'USO

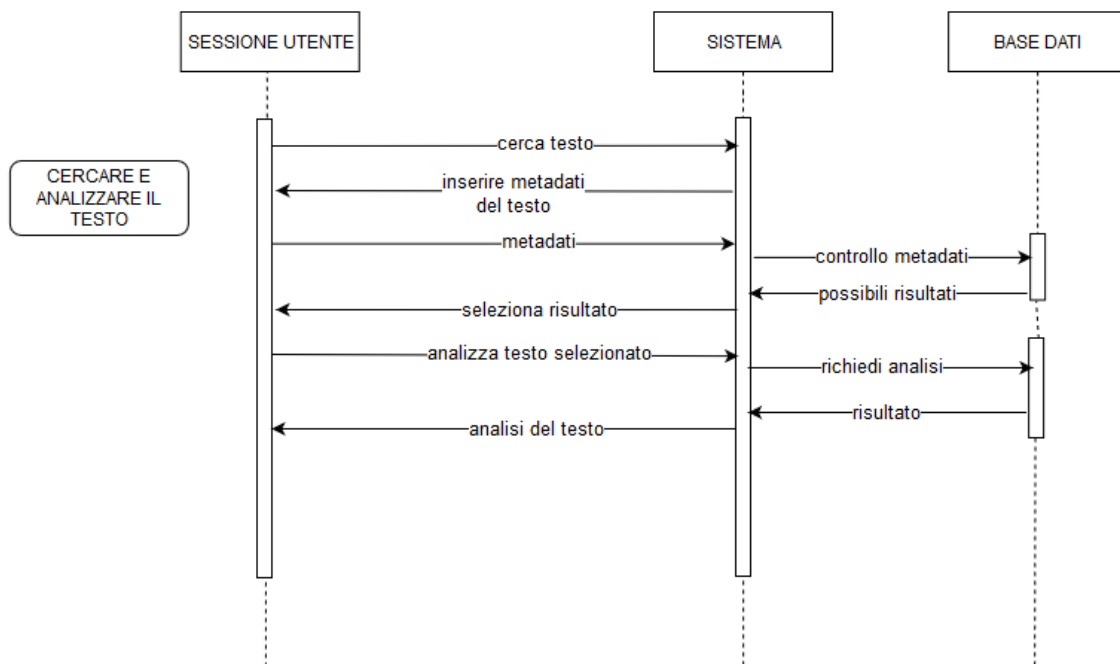


3.4 DIAGRAMMA DELLE ATTIVITÀ

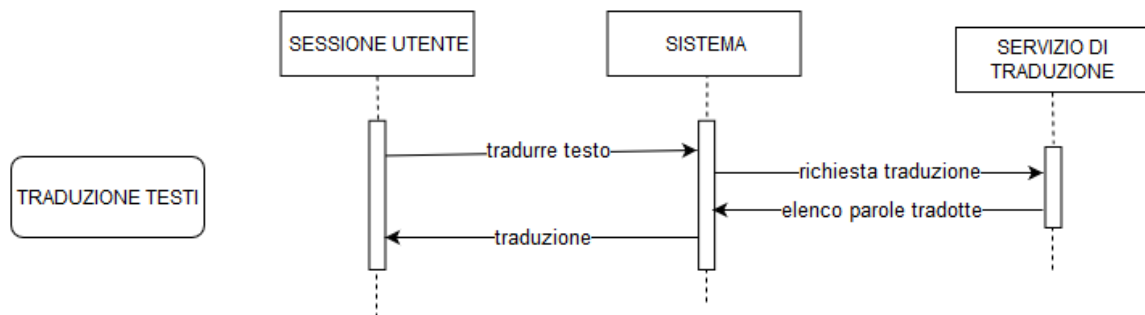




3.5.4 Cerca e analizza il testo

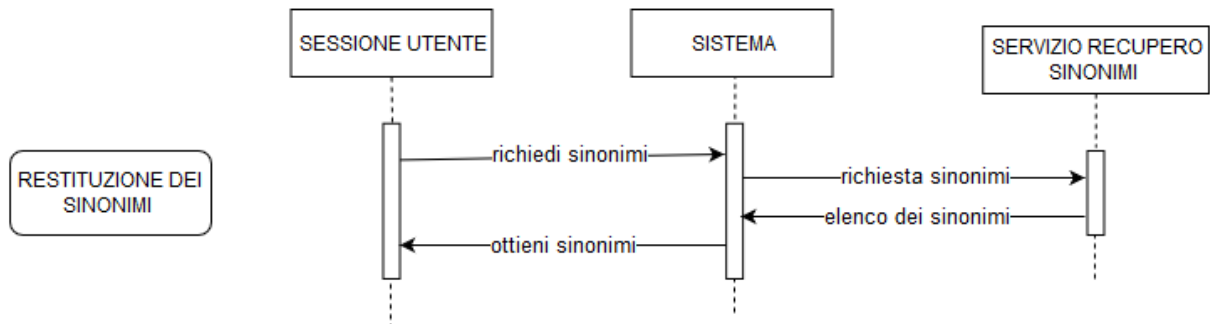


3.5.5 Traduzione dei testi



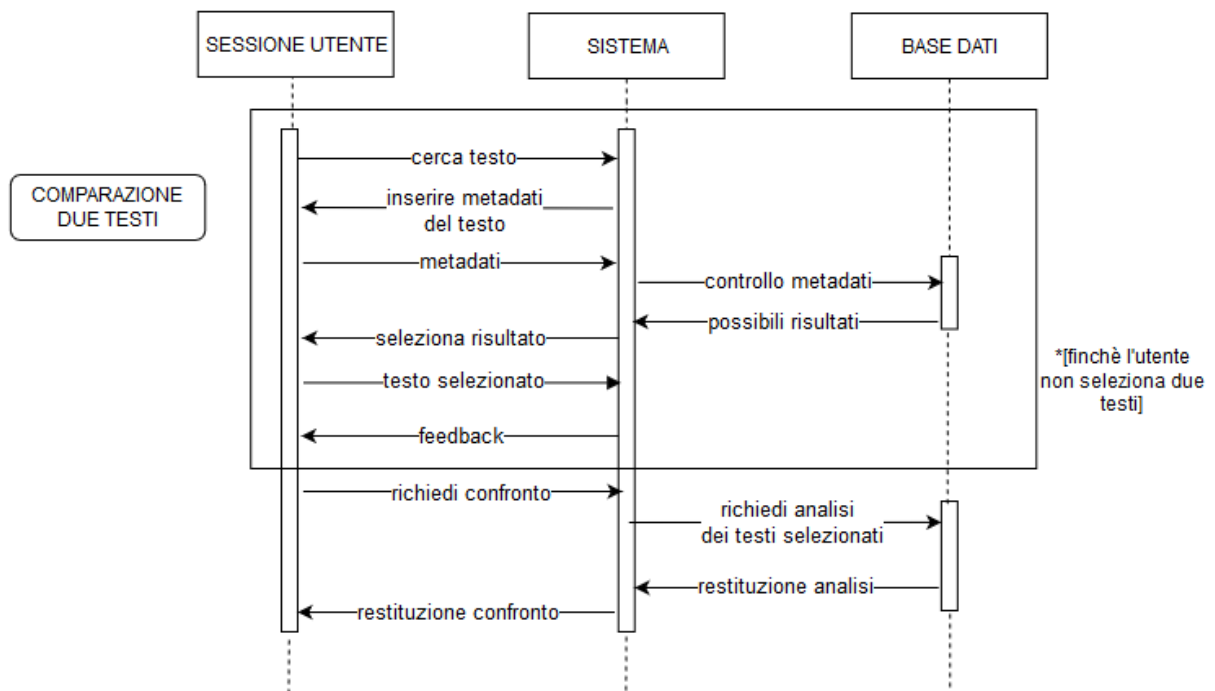
*I termini sono già caricati nel database

3.5.6 Restituzione sinonimi

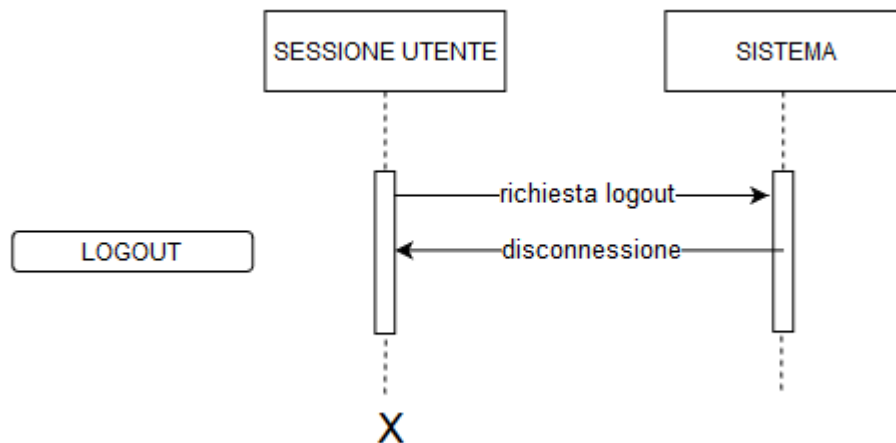


*I termini sono già caricati nel database

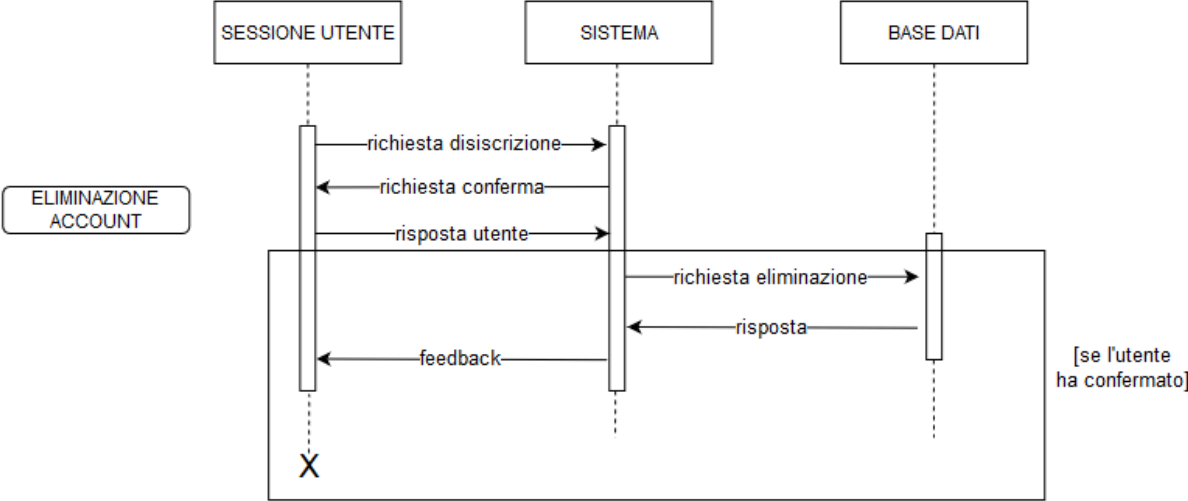
3.5.7 Comparare due testi



3.5.8 Logout



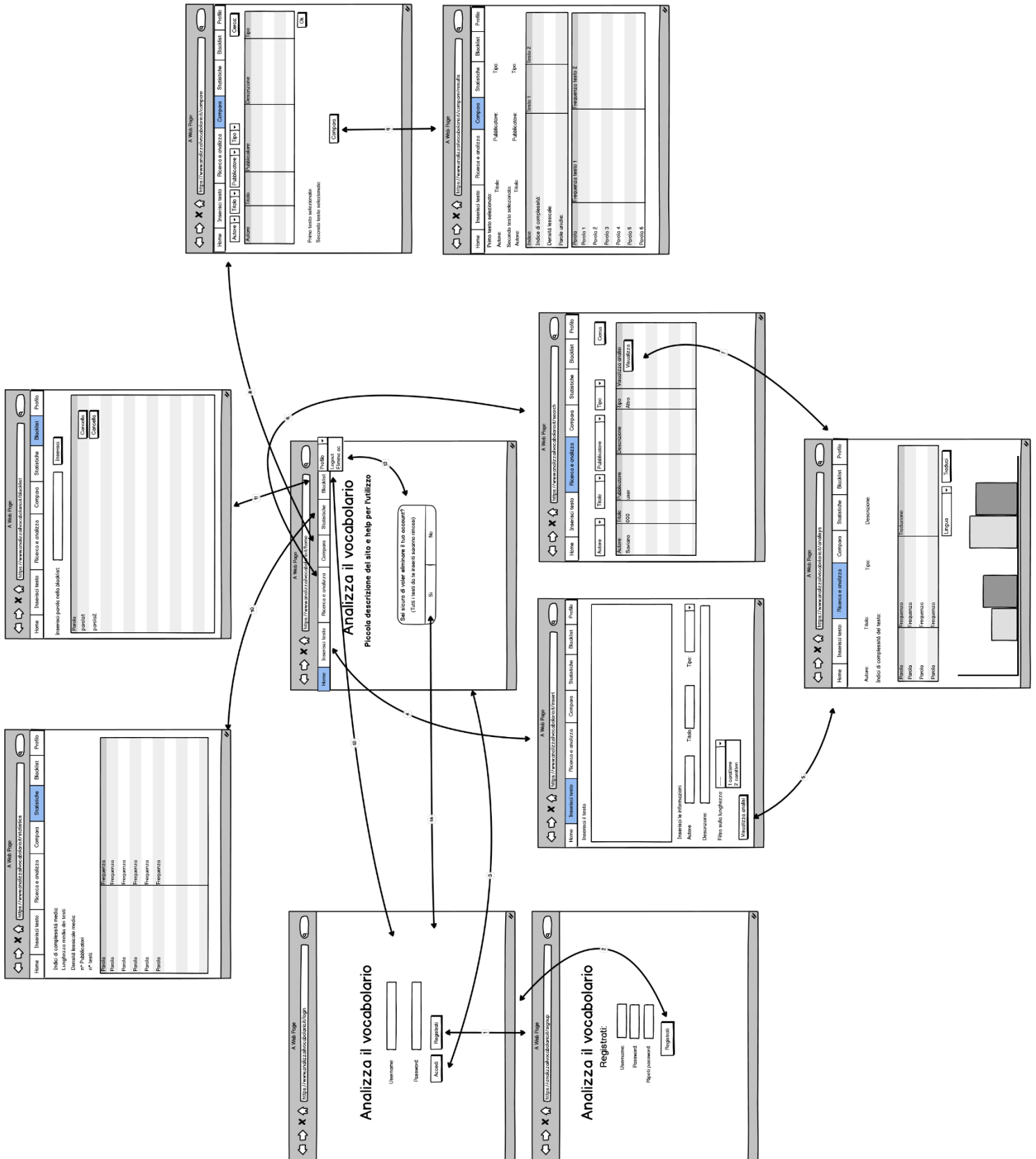
3.5.9 Elimina account



4 USER STORIES

IMPORTANTISSIME	IMPORTANTI	MENO IMPORTANTI	POSSIBILI AGGIUNTE
Come utente registrato voglio inserire testi nella piattaforma per analizzarli. Story Point: 14	Come utente voglio potermi registrare alla comunità per poter accedere ai servizi offerti dal sito. Story Point: 4	Come utente registrato voglio la possibilità di fare un logout per uscire dalla sessione. Story Point: 5	Come utente registrato voglio visualizzare il numero di parole totali di un testo. Story Point: 23
Come utente registrato voglio poter ricercare testi presenti nella piattaforma Story Point: 20	Come utente voglio effettuare il login per usufruire dei servizi del sito. Story Point: 5	Come utente registrato voglio avere una sorta di guida alle funzionalità che posso effettuare (una sorta di help). Story Point: 2	Come utente registrato voglio visualizzare il numero di parole uniche del testo. Story Point: 23
Come utente registrato voglio visualizzare la frequenza delle parole di un testo per studiare le caratteristiche del documento. Story Point: 23	Come utente registrato vorrei poter visualizzare le traduzioni dei termini per imparare le corrispettive traduzioni. Story Point: 33	Come utente voglio poter eliminare il mio account. Story Point: 8	Come utente registrato voglio visualizzare la densità lessicale di un testo. Story Point: 23
Come utente registrato voglio visualizzare l'indice di complessità di un testo per studiarne la complessità. Story Point: 40			Come utente della comunità voglio poter visualizzare sinonimi per arricchire il mio linguaggio. Story Point: 90
Come utente della comunità voglio poter confrontare due testi per mostrare differenze di frequenze e complessità. Story Point: 60			

5 MOCKUP



*È stata allegata una versione a dimensione reale del mockup.

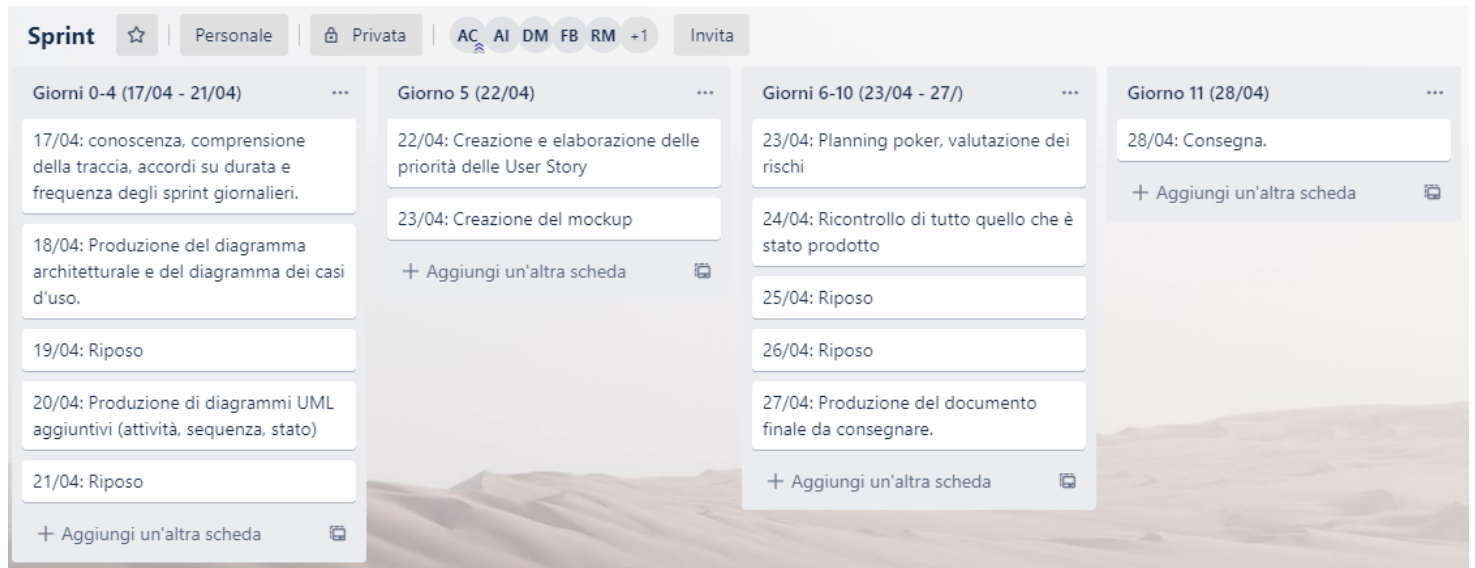
5.1 DESCRIZIONE MOCKUP

1. La prima pagina del sito web che si presenta è quella del LOG IN
2. Se l'utente non è ancora registrato allora attraverso il bottone "REGISTRATI" verrà reindirizzato alla pagina di registrazione, dove dovrà inserire i propri dati. Una volta registrato dovrà effettuare l'accesso.
3. Dopo aver effettuato l'accesso al sito l'utente si troverà nella pagina "HOMEPAGE" dove potrà leggere le "istruzioni d'uso" e una breve descrizione del sistema "Analizza Vocabolario". All'interno delle pagine del sito visibili solo agli utenti loggati è presente una navbar che permetterà all'utente di navigare nel sito e usufruire dei servizi offerti.
4. Attraverso il bottone "INSERISCI TESTO" nella navbar l'utente si troverà in una pagina in cui potrà incollare il testo desiderato e inserire i suoi metadati per poi pubblicarlo.
5. Se la checkbox "VISUALIZZA RISULTATI" sarà attivata, l'utente sarà reindirizzato in una pagina contenente i risultati dell'analisi del testo appena pubblicato. Inoltre, egli potrà visualizzare, opzionalmente, la traduzione, nella lingua desiderata, delle parole contenute nel suo testo.
6. Cliccando sul bottone "RICERCA E ANALIZZA" l'utente potrà ricercare un testo attraverso i suoi metadati (in base alle selezioni delle combobox) e analizzarlo.
7. L'analisi del testo ricercato verrà mostrata in una pagina in cui compariranno varie statistiche relative alle parole. Inoltre, egli potrà visualizzare, opzionalmente, la traduzione, nella lingua desiderata, delle parole contenute nel suo testo.
8. Selezionando il bottone "COMPARA" sarà possibile comparare due testi, ricercandoli attraverso i loro metadati (in base alle selezioni delle combobox) uno alla volta. La comparazione riguarderà la visualizzazione dei confronti tra le statistiche delle parole contenute nei due testi.
9. Un utente potrà uscire dalla propria sessione attraverso il bottone "LOGOUT" in ogni momento.
10. Un utente, nella pagina "HOMEPAGE", potrà decidere di eliminare il proprio account attraverso il bottone "ELIMINA ACCOUNT" nel footer della pagina. Attraverso una notifica pop-up verrà richiesta la conferma di eliminazione.

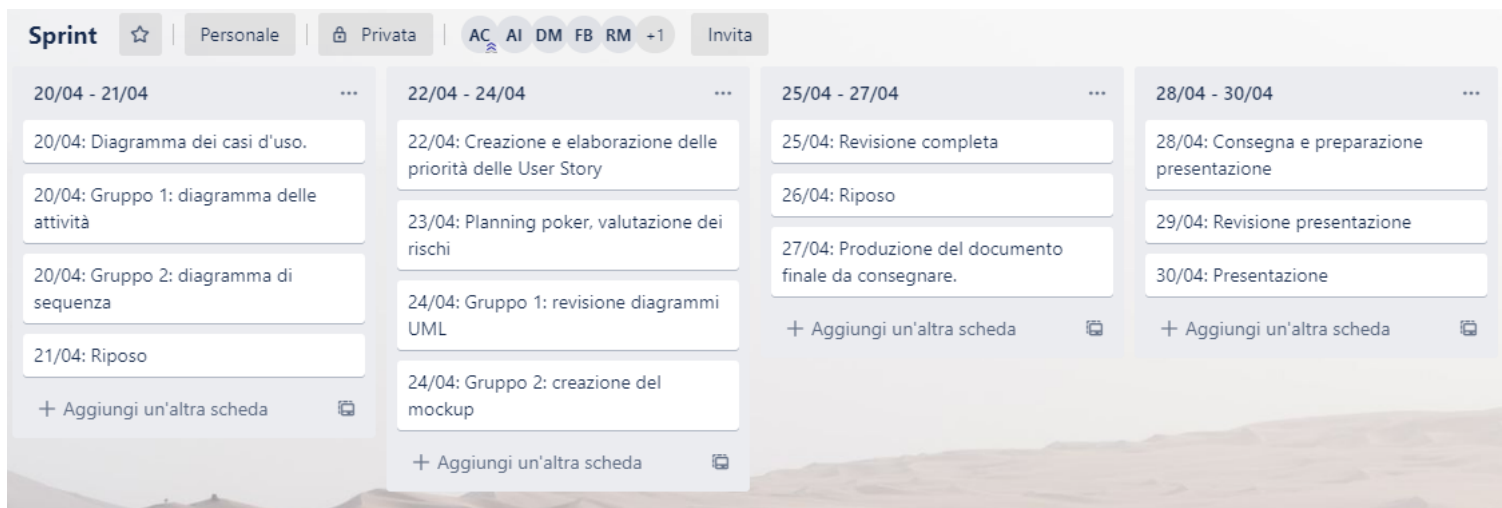
6 CRITICITÀ

RISCHIO	IMPATTO	PROBABILITA'	TOTALE
Effettuare la registrazione di un utente	10	2	20
Effettuare il login al sito	10	2	20
Effettuare il logout dal sito	3	1	3
Inserire un testo e i relativi metadati	8	3	24
Utilizzo dei servizi del sito solo per gli utenti registrati	5	3	15
Calcolare l'indice di complessità	10	7	70
Ricerca un testo nel sito	9	7	63
Comparare due testi	10	8	80
Memorizzare un testo e i relativi metadati nel database	10	3	30
Recuperare la traduzione di una parola	5	7	35
Calcolare la frequenza delle parole di un testo	10	4	40
Creazione e aggiornamento del database	10	2	20
Visualizzare i risultati dell'analisi	10	1	10
Eliminazione account	2	4	8
Eliminazione testi associati all'account eliminato	4	6	24
Preprocessing del testo	8	2	16
RISCHIO USER STORIES FACOLTATIVE	IMPATTO	PROBABILITA'	TOTALE
Calcolare numero di parole uniche del testo	1	4	4
Calcolare numero di parole totali del testo	1	4	4
Calcolare la densità lessicale	1	4	4
Recuperare i sinonimi	3	8	24

7 DOCUMENTAZIONE SPRINT



Questo è stato il risultato della nostra prima pianificazione; questo sprint ha ricevuto una modifica nel giorno 3 per via di un ritardo nella creazione del diagramma dei casi d'uso successa nel giorno 1. Il nuovo sprint è stato il seguente:



Il ritardo ha fatto sì che tutte le attività fossero scalate di un giorno, facendo arrivare la revisione completa al 25/04. Oltre a ciò abbiamo deciso di modificare le modalità e l'ordine di alcune attività; in particolare abbiamo deciso:

- di dividerci nella produzione dei diagrammi UML aggiuntivi (quello di attività e quello di sequenza);
- di creare le user story e di darle una priorità insieme (inizialmente avevamo previsto che solo una piccola parte del gruppo si sarebbe occupato di questo punto);
- di spostare la creazione del mockup alla fine del progetto.