



PEBKAC

Gruppo: 11

Email: pebkacswe@gmail.com

Docs: <https://pebkac-swe-group-11.github.io>

GitHub: <https://github.com/PEBKAC-SWE-Group-11>



Università degli Studi di Padova

Corso di Laurea: Informatica

Corso: Ingegneria del Software

Anno Accademico: 2024/2025

Verbale Esterno

10 dicembre 2024

Informazioni sul documento:

Responsabile	Matteo Gerardin
Verificatore	Tommaso Zocche
Redattore	Matteo Gerardin
Uso	Esterno
Destinatari	Vimar S.p.A. Tullio Vardanega Riccardo Cardin

Abstract:

Seconda riunione SAL del gruppo con Vimar S.p.A. Sono stati discussi e chiariti i dubbi e le perplessità relativi alle tecnologie studiate.

Registro delle versioni

Versione	Data	Autore	Ruolo	Descrizione
1.0.0	16/12/24	Alessandro Benin	Responsabile	Approvazione e rilascio
0.1.0	13/12/2024	Tommaso Zocche	Verificatore	Verifica
0.0.1	11/12/2024	Matteo Gerardin	Responsabile	Stesura iniziale

Indice

1	Informazioni generali	4
2	Riassunto della riunione	5
3	Todo	9

1 Informazioni generali

- **Tipo riunione:** esterna
- **Luogo:** telematica, Teams
- **Data:** 10 dicembre 2024
- **Ora inizio:** 15.00
- **Ora fine:** 16.20
- **Presenti:**
 - Alessandro Benin
 - Ion Bourosu
 - Matteo Gerardin
 - Derek Gusatto
 - Davide Martinelli
 - Matteo Piron
 - Tommaso Zocche
 - ★ Mariano Sciacco (Vimar S.p.A.)
 - ★ Francesca Stival (Vimar S.p.A.)
- **Assenti:**

2 Riassunto della riunione

Come previsto, si è proceduto seguendo l'ordine del giorno previsto per le riunioni SAL:

1. **Presentazione dei ruoli:** vengono elencati i ruoli assunti dai membri del gruppo in questo periodo:
 - Alessandro Benin - Progettista;
 - Ion Bourosu - Analista;
 - Matteo Gerardin - Responsabile;
 - Derek Gusatto - Analista;
 - Davide Martinelli - Verificatore;
 - Matteo Piron - Amministratore;
 - Tommaso Zocche - Analista.
2. **Panoramica ad ampio spettro:** sono stati discussi gli avanzamenti effettuati sulla documentazione, prestando particolare attenzione ai casi d'uso, presenti nell'Analisi dei Requisiti, ed inoltre è stato effettuato un confronto con i rappresentanti dell'azienda proponente relativamente a quanto emerso dallo studio delle tecnologie.
3. **Attività completate ed in corso:** sono state esposte le attività che sono state svolte in questo periodo, suddividendole principalmente in stesura della documentazione e ricerca tecnologica.

Le prime riguardano i progressi effettuati con la stesura di:

- Piano di Progetto;
- Norme di Progetto;
- Glossario;
- Analisi dei Requisiti.

Inoltre, in questo periodo, è stata iniziata la redazione del Piano di Qualifica. Nonostante gli avanzamenti raggiunti, queste attività sono tuttora in corso, in quanto deve essere ancora terminata la prima stesura di tutti i documenti sopraelencati. Ci si è soffermati particolarmente sull'Analisi dei Requisiti, ed in particolare sui casi d'uso, che sono stati presi nuovamente in esame dopo che sono emerse ulteriori informazioni che li riguardano, e per cui si sono discusse ulteriori modifiche che sarà necessario applicare.

Le seconde invece riguardano lo studio e la sperimentazione pratica delle tecnologie necessarie alla realizzazione del progetto. Nello specifico sono state oggetto di queste attività:

- Flask, Angular e Vue.js per la realizzazione del frontend;
- PostgreSQL con l'estensione pgvector (per la creazione degli indici vettoriali) per la realizzazione del database vettoriale;
- Scrapy e OCRmyPDF per il web scraping;
- Python per la realizzazione del backend;

- Llama 3.1, Mistral, Bert e Phi come modello LLM.

Oltre ad essere state analizzate e testate individualmente, si è anche provato ad integrare alcune di esse.

4. **Prossime attività da svolgere:** nel prossimo periodo si è deciso di:

- Effettuare un'analisi e delle prove con un nuovo modello LLM proposto, ovvero OpenGPT-x;
- Effettuare delle prove di interrogazione dei modelli LLM utilizzando dei dataset presi da della documentazione di Vimar;
- Studiare ed effettuare delle prove con Docker per la progettazione e la realizzazione dell'architettura e realizzare un primo schema di quest'ultima;
- Effettuare ulteriori prove per le operazioni di scraping e di embedding per le immagini;
- Effettuare la redazione di un verbale interno in cui vengono elencate le scelte prese in merito alle tecnologie ed in cui vengono spiegate e motivate queste scelte;
- Proseguire con le prove di integrazione fra le varie tecnologie, in direzione della realizzazione del PoC;
- Proseguire con la redazione di Piano di Progetto, Norme di Progetto, Glossario, Analisi dei Requisiti e Piano di Qualifica.

5. **Discussione di dubbi e domande:** vengono riportate di seguito le domande che sono state poste durante la riunione con le relative risposte:

A) **Riguardo la scelta della tecnologia per il frontend, dopo un confronto effettuato è risulta che Flask ed Angular svolgono due funzioni sostanzialmente diverse. È possibile utilizzare Flask insieme ad Angular? Si può scegliere solamente uno dei due scartando l'altro in tranquillità?**

Mentre Angular è un linguaggio che viene utilizzato totalmente a livello frontend, Flask può essere utilizzato sia per realizzare delle API (quindi per costruire sostanzialmente un server di backend che fornisca delle risposte) sia per renderizzare una pagina web. Quindi è possibile utilizzare Angular che avrà poi bisogno di essere collegato tramite delle API, che possono essere realizzate anche con lo stesso Flask, oppure è possibile utilizzare direttamente Flask per la renderizzazione della pagina, senza doversi successivamente preoccupare di gestire delle API per realizzare questo tipo di implementazione.

B) **Perché si dovrebbe preferire Angular a Vue.js?**

Angular è un linguaggio che esiste da parecchi anni e che ha subito svariati aggiornamenti, ed è molto conosciuto dalla community, quindi in un futuro potrebbe essere utilizzato nel curriculum. Inoltre possiede una serie di pattern che sono in grado di semplificare il lavoro, al costo di saperli utilizzare. Infine è un linguaggio molto rigoroso. Vue.js, invece, è un linguaggio molto giovane che risulta essere particolarmente complesso per certi aspetti, ma è un'ottima alternativa se si vuole evitare la rigidità di Angular.

C) Esiste un modo per verificare se le risposte fornite dall'LLM sono corrette e conformi alla documentazione presa dal sito di Vimar?

È possibile utilizzare due diverse strategie:

- È possibile caricare il pdf di un manuale su GPT e chiedere delle possibili domande con le relative risposte, andando poi a porre le stesse domande al nostro modello LLM e confrontando le sue risposte con quelle fornite da GPT;
- È possibile porre direttamente una domanda relativa alla documentazione su GPT, che dovrebbe essere in grado di rispondere dato che si tratta di documenti pubblici, e, successivamente, è possibile porre la stessa domanda al nostro modello LLM e confrontare le due risposte.

D) Nella fase di web scraping è necessario estrarre tutte le informazioni disponibili nella documentazione? Ed è necessario farlo per tutte le lingue disponibili o solamente per una o due?

Alcune delle informazioni presenti nella documentazione sono ripetute in molteplici manuali, esattamente nello stesso modo, quindi sarà necessario ottimizzare il database per gestire queste ridondanze. Inoltre, relativamente alle versioni dei manuali in differenti lingue, è necessario estrarre solamente tre diverse, dato che le traduzioni sono letterali e non si rischia di avere perdita di informazione.

E) Come è possibile effettuare l'operazione di scraping delle immagini presenti nella documentazione? Un'alternativa che abbiamo considerato è stata generare un'alternativa testuale tramite un modello LLM che esprima il contenuto informativo dell'immagine? È una soluzione valida?

Può essere una soluzione assolutamente valida, ma necessiterà di una pipeline automatizzata, in quanto effettuare questo procedimento manualmente per tutte le immagini richiederebbe un quantitativo enorme di tempo. Un'altra alternativa che potrebbe essere valutata è "fare una foto" all'intera pagina del manuale e trovare un modo per far sì che il modello LLM riconosca che in essa è presente un'immagine, permettendone così l'estrazione.

F) È necessario implementare l'embedding delle immagini da parte del modello LLM per renderle disponibili nel database e poi poterle utilizzare?

Per poter effettuare questa operazione è necessario che il modello sia multimodale, ovvero che sia in grado di comprendere sia testo che immagini, e potenzialmente anche audio. Sarebbe inoltre necessario capire come elaborare il tutto in maniera ottimale. Un'alternativa che può essere utilizzata per realizzare questo obiettivo è, una volta realizzata completamente la parte di embedding nel database, quando si passano gli indici che contengono le informazioni necessarie al modello per fornire la risposta, si può fare in modo che recuperi anche il link dell'immagine che "potrebbe tornare utile" all'utente che ha posto una determinata domanda, senza che il modello sappia effettivamente cosa rappresenta l'immagine in questione.

G) Se il modello LLM possiede la capacità di comprendere le immagini molto bene si può evitare di realizzare l'alternativa testuale che era stata discussa precedentemente?

In questo caso si potrebbe evitare di creare l'alternativa testuale, ma sarebbe più complesso gestire il reperimento delle immagini nel database. Una possibilità per la gestione di questa problematica potrebbe essere inserire in linguaggio naturale un contesto iniziale da passare al modello per istruirlo a svolgere questa operazione di ricerca dell'immagine oltre alle informazioni testuali. Viene comunque consigliato di analizzare per gradi il problema dell'embedding delle immagini e di trascurarlo per il momento in favore dell'embedding del testo, ritenuto più importante e semplice.

H) È giusto considerare come attori anche degli elementi interni al sistema?

Un attore può essere definito come chiunque o qualunque parte del sistema che abbia delle necessità che devono essere in qualche modo sopperite per raggiungere un obiettivo. In virtù di questa definizione, quindi, possono essere considerati attori anche componenti interni al sistema, come ad esempio il modello LLM, nonostante questa scelta debba essere chiaramente motivata.

I) È possibile utilizzare dei diagrammi di attività invece dei diagrammi di casi d'uso per rappresentare i casi d'uso che riguardano gli attori interni al sistema?

Sì, è possibile in quanto, grazie alle regole a cui questo diagramma fa riferimento, è possibile capire chiaramente il flusso delle attività che si verificano e di quelle che non si verificano.

È stato inoltre chiesto di spostare la data per la prossima riunione SAL dal 24/12/2024 al 23/12/2024 con orario da decidere da parte del gruppo, e che verrà comunicato ai referenti dell'azienda tramite email.

3 Todo

Durante la riunione sono emersi i seguenti task da svolgere.

Assegnatario	Task Todo
Matteo Gerardin	Stesura Verbale Esterno 10/12/2024
Derek Gusatto	Progettazione e realizzazione di uno schema dell'architettura
Matteo Piron	Stesura del verbale interno in cui vengono motivate le scelte tecnologiche
<i>autoassegnazione</i>	Prosecuzione della redazione della documentazione

Task non produttivi:

Assegnatario	Task Todo
<i>autoassegnazione</i>	Studio e prove pratiche di OpenGPT-x
Derek Gusatto, Matteo Piron	Studio e prove pratiche di Docker
Matteo Piron	Prove di interrogazione dei modelli LLM con dataset ricavati dalla documentazione di Vimar
Tommaso Zocche	Continuazione delle prove di scraping e di trascrizione dei PDF
Matteo Piron	Prove di integrazione tra le varie tecnologie selezionate

Firma del referente Vimar S.p.A.: _____