



7zpus.swe@gmail.com

Verbale Esterno Gruppo di Progetto e Sanmarco Informatica

Data: 2026/1/22

Durata: 15 minuti

Luogo: incontro online (Meet)

Tabella di Versionamento

Versione	Data	Autore	Verificatore	Descrizione
0.1	2026/1/22	Rocco Matteo A.	Fattoni Antonio	Creazione del verbale e stesura iniziale

Indice

1	Ordine del Giorno	2
2	Organizzazione incontri successivi	2
3	Svolgimento e Discussione	2
3.1	Esposizione dello stato di avanzamento del PoC	2
3.2	Ricerca semantica	3
4	Domande e Risposte	3
5	Conclusione	3

Partecipanti

- Fattoni Antonio
- Georgescu Diana
- Gingilino Aaron
- Laoud Zakaria
- Rocco Matteo Alberto
- Soligo Lorenzo
- Vigolo Davide
- Rossi Andrea
- Sarra Luca

1 Ordine del Giorno

1. Esposizione dello stato di avanzamento del PoC.

2 Organizzazione incontri successivi

Sono state programmate le date degli incontri successivi, che si terranno ogni 2 settimane, fino al 16/04/2026, data in cui è prevista la consegna della Product Baseline.

3 Svolgimento e Discussione

3.1 Esposizione dello stato di avanzamento del PoC

Durante l'incontro il Team ha esposto lo stato di avanzamento del Proof of Concept (PoC), in particolare ha riportato i diversi approcci all'indicizzazione e alla ricerca del DIP che sono stati considerati:

- indicizzazione pura con database SQL.
- indicizzazione mista con file json e database SQL.
- indicizzazione basata sulla libreria Tantivy.

Il primo approccio è stato illustrato brevemente in quanto ancora in fase di sviluppo e non ancora testato.

Il secondo approccio, basato sulla prima versione di PoC presentata alla proponente, è stato migliorato dal punto di vista dell'usabilità e corretto per quanto riguarda la verifica dell'integrità dei file tramite funzione di hash in Base64.

Il terzo approccio, basato sulla libreria Tantivy, sebbene interessante, è stato scartato in quanto sviluppata in Rust, cosa che avrebbe richiesto il mantenimento continuo del codice per adattare il PoC alle nuove versioni del linguaggio e della libreria stessa.

Le prime due versioni di PoC verranno testate e confrontate per valutare quale delle due implementazioni risulti più efficiente, in tal senso sarà significativo confrontare

i tempi di indicizzazione del pacchetto DIP di test fornito in data 15/01/2026 dalla proponente su richiesta del Team, che ha una dimensione di diversi ordini di grandezza superiore rispetto ai pacchetti DIP utilizzati finora.

3.2 Ricerca semantica

Le risorse previste per lo studio della terzo approccio sono state reindirizzate allo studio e alla progettazione dell'implementazione della ricerca semantica, che è un requisito opzionale ma che il Team ritiene importante implementare già nel PoC, in quanto rappresenta una funzionalità chiave del prodotto finale e potrebbe essere complessa da implementare.

In particolare è stata individuata una libreria integrata in SQLite in linguaggio C, `sql-vec`, che permette di effettuare ricerche semantiche basate su vettori, che sembra essere adatta alle esigenze del progetto come alternativa alla libreria FAISS.

4 Domande e Risposte

- **Domanda:** È desiderabile implementare la ricerca semantica non solo sui metadati, ma anche sul contenuto dei documenti?
- **Risposta:** Non riteniamo necessario implementare la ricerca semantica sul contenuto dei documenti, in quanto comporterebbe la trasformazione di tutti i documenti in vettori, diventando un lavoro particolarmente oneroso.
Per quanto riguarda la traduzione in vettori dei metadati, esistono diversi modelli di embedding che operano in locale, ad esempio con Docker Model Runner oppure ancora modelli del tool Ollama.

5 Conclusione

L'incontro si è concluso con una breve presentazione del PoC e si rimanda alla prossima riunione per la presentazione dei risultati dei test sulle prime due versioni di PoC e per la discussione dell'implementazione della ricerca semantica.

Data: _____

Firma: _____
[Sanmarco Informatica
SPA]