

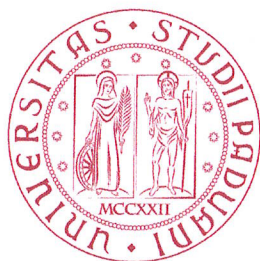


Verbale Riunione 2024-05-06

Gruppo Argo — Progetto ChatSQL

Informazioni sul documento

Versione	1.0.0
Approvazione	Raul Pianon
Uso	Esterno
Distribuzione	Zucchetti S.p.A. Prof. Tullio Vardanega Prof. Riccardo Cardin Gruppo Argo



**UNIVERSITÀ
DEGLI STUDI
DI PADOVA**

Registro delle modifiche

Ver.	Data	Redazione	Verifica	Descrizione
1.0.0	2024-05-28	Raul Pianon	Riccardo Cavalli, Sebastiano Lewental	Approvazione documento
0.0.1	2024-05-10	Marco Cristo	Riccardo Cavalli, Sebastiano Lewental	Stesura del documento

Indice

1	Informazioni	3
1.1	Descrizione	3
1.2	Partecipanti	3
1.3	Glossario	3
2	Riunione	4
2.1	Ordine del giorno	4
2.2	Argomenti e temi dell'incontro	4
2.2.1	Tecnologie per lo sviluppo della web app	4
2.2.2	Resoconto sullo studio preliminare di txtai	4
2.2.3	Analisi dei casi d'uso	5
2.2.4	Organizzazione della presentazione su Streamlit	5
3	Todo / In Progress	6

1 Informazioni

- **Inizio incontro:** 17:30
- **Fine incontro:** 18:15
- **Pianificazione incontro:** Mail
- **Tipo incontro:** remoto (Zoom)

1.1 Descrizione

Questo verbale riporta gli aggiornamenti riguardanti lo studio e il collaudo delle tecnologie che il team intende adottare, assieme ad alcune domande poste alla *Proponente*_e relativamente allo sviluppo dell'applicativo, in vista della realizzazione del *PoC*_e.

1.2 Partecipanti

- **Argo:**
 - Tommaso Stocco → 45 minuti
 - Marco Cristo → 45 minuti
 - Riccardo Cavalli → 45 minuti
 - Raul Pianon → 45 minuti
 - Martina Dall'Amico → 45 minuti
 - Sebastiano Lewental → 45 minuti
- **Zucchetti S.p.A.:**
 - Gregorio Piccoli

1.3 Glossario

Allo scopo di evitare incomprensioni relative al linguaggio utilizzato nella documentazione di progetto, viene fornito un *Glossario*, nel quale ciascun termine è corredato da una spiegazione che mira a disambiguare il suo significato. I termini tecnici, gli acronimi e i vocaboli ritenuti ambigui vengono formattati in corsivo all'interno dei rispettivi documenti e marcati con una lettera _e in pedice. In questo documento viene formattata solamente la prima ricorrenza di un termine definito nel *Glossario*.



2 Riunione

2.1 Ordine del giorno

- Discutere delle tecnologie per lo sviluppo dell'applicativo;
- Presentare i dubbi emersi durante lo studio di *txtai*_e;
- Valutare l'ampliamento dei casi d'uso_e.

2.2 Argomenti e temi dell'incontro

2.2.1 Tecnologie per lo sviluppo della web app

Domanda: Tra le diverse tecnologie solitamente indicate per lo sviluppo di un'applicazione web (quali Django, Flask, React, Vue.js), abbiamo individuato il *framework*_e *Streamlit*_e. Il framework in questione permette di mantenere il linguaggio *Python*_e per sviluppare l'interfaccia grafica, seppur conceda meno spazio alla personalizzazione. Secondo lei, potrebbe essere una scelta opportuna?

Risposta: La Proponente non è familiare col framework *Streamlit*, ma ne approverebbe l'utilizzo a seguito di una presentazione informativa da parte del gruppo in cui si spiegano le motivazioni di tale scelta.

2.2.2 Resoconto sullo studio preliminare di *txtai*

Descrizione: Il programmatore espone i risultati conseguiti durante lo studio della libreria *txtai*_e, con particolare enfasi sulla componente di *indicizzazione*_e e ricerca semantica. Pone poi l'attenzione sulla necessità o meno delle fasi di pre-processing della richiesta dell'utente, nel tentativo di semplificare la ricerca semantica. L'idea del team sarebbe di processare la richiesta in linguaggio naturale per *tokenizzare*_e solo le componenti semantiche ritenute fondamentali. Allo stesso modo, sarebbe prevista un'elaborazione del *dizionario dati*_e in formato *JSON*_e per renderlo simile al linguaggio naturale, cosicché i *modelli*_e, allenati su quest'ultimo, riescano ad estrarne con più facilità il significato.

Risposta: La Proponente ritiene che le fasi di pre-processing non siano necessarie, poiché il formato del dizionario dati risulta sufficiente per creare degli *embeddings*_e efficaci. Illustra poi l'utilizzo di *txtai* per la ricerca semantica tramite *sentence-BERT*_{ino} di efederici, un modello operante su lingua italiana. Inoltre, la Proponente invita a trarre vantaggio dalle funzioni di load e save degli indici, per ridurre l'*overhead*_e e velocizzare la fase di testing.

Descrizione: Il programmatore espone il problema dello scarso punteggio dei risultati restituiti dalla funzione *search* di *txtai*, e suggerisce la mitigazione del problema tramite l'aggiunta di un indice intermedio.

Risposta: La Proponente accoglie la proposta e ne invita la sperimentazione. Riguardo allo score basso degli output forniti dalla *search*, la Proponente suggerisce l'esplorazione di più modelli. Spiega inoltre che *txtai* utilizza, in maniera del tutto trasparente, la libreria *FAISS*_e, che si occupa della *vettorizzazione*_e e *indicizzazione* oltre

che della ricerca semantica. In aggiunta, la Proponente suggerisce di confrontare FAISS con altre librerie.

2.2.3 Analisi dei casi d'uso

Domanda: L'analista illustra i *casi d'uso*₆ relativi alla funzionalità di *debug*₆ disponibile per il profilo Tecnico. Il report ideato dal team comprende una lista di tabelle con annesso il punteggio di rilevanza. La Proponente immagina delle funzionalità aggiuntive per il debugging?

Risposta: La Proponente conferma come lo scopo del debugging sia mostrare l'accuratezza e la completezza del dizionario dati caricato dal Tecnico. Suggerisce poi che venga fornita una spiegazione sull'eventuale esclusione di un risultato pertinente a favore di uno meno attinente.

Descrizione: L'analista chiede se sia ragionevole generare un elenco di punteggi, mostrando l'intero *indice*₆ o gran parte di esso.

Risposta: La Proponente spiega che grazie alla logica di *back-end*₆, il processo di ricerca non avviene per calcolo sequenziale, ma attraversa una struttura gerarchica multistrato. Pertanto, la query utente non deve essere confrontata con tutti i vettori dell'indice. Eseguire la funzione search non è costoso, mentre lo sarebbe visitare in maniera sequenziale l'intero indice. L'unica operazione onerosa, per cui potrebbero essere necessari anche dei minuti, riguarda la creazione dell'indice; per tale motivo sono state introdotte le già menzionate funzioni load e save.

2.2.4 Organizzazione della presentazione su Streamlit

Descrizione: Il responsabile domanda alla Proponente se preferisca una presentazione del framework Streamlit, evidenziandone vantaggi e svantaggi, o se preferisca una relazione da spedire via email.

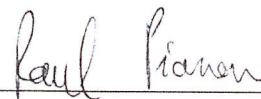
Risposta: La Proponente suggerisce al team di elaborare una breve presentazione da esporre durante l'incontro successivo.

3 Todo / In Progress

Ticket	Incarico	Incaricato/a	Scadenza
ARGO-54	Stesura verbale riunione	Marco Cristo	2024-05-10
ARGO-37	Analisi dei pro e contro di Streamlit ed elaborazione di una presentazione	Riccardo Cavalli, Sebastiano Lewental	2024-05-15
ARGO-52	Valutazione delle alternative a Streamlit (Flask, Django, React e Vue.js)	Sebastiano Lewental, Martina Dall'Amico, Raul Pianon	2024-05-20
ARGO-48	Benchmark _e dei modelli	Riccardo Cavalli, Raul Pianon, Marco Cristo	2024-05-15

Luogo e Data:
Padova (PD) 2024-05-06

Firma: _____



Responsabile: Raul Pianon

Per approvazione:

Firma: _____



Referente: Gregorio Piccoli (Zucchetti S.p.A.)

Zucchetti S.p.A.

Via Solferino, 1 - 26900 LODI

Tel. 0371.5945700 - Fax 0371.5945753

Sede Op.: Via G. Cittadella, 7 - 35137 PADOVA

P. IVA e Cod. Fisc. 05006900962

Verbale Riunione 2024-05-06

📄 v 1.0.0

Marchetti S.p.A.
Via Solferino, 1 - 20090 LODI
Tel. 0371.8945700 - Fax 0371.8945283
Sede Op. Via G. Garibaldi, 7 - 37131 PADOVA
P. IVA e Cod. Fisc. 02009000282