



ARGO

Verbale Riunione 2024-05-22

Gruppo Argo — Progetto ChatSQL

Informazioni sul documento

Versione	1.0.0
Approvazione	Martina Dall'Amico
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
0.0.1	2024-06-09	Martina Dall'Amico, Riccardo Cavalli	Stesura del documento	

Indice

1 Informazioni

- **Inizio incontro:** 14:30
- **Fine incontro:** 15:30
- **Pianificazione incontro:** Mail
- **Tipo incontro:** remoto (Zoom)

1.1 Descrizione

Durante il meeting, il gruppo ha presentato alla *Proponente*_e le caratteristiche delle tecnologie individuate per sviluppare l'applicazione. Il discorso si è poi concentrato su una discussione più approfondita di casi d'uso e metriche di qualità e sulla risoluzione dei dubbi del team.

1.2 Partecipanti

- **Argo:**
 - Tommaso Stocco → 1 ora
 - Marco Cristo → 1 ora
 - Riccardo Cavalli → 1 ora
 - Raul Pianon → 1 ora
 - Martina Dall'Amico → 1 ora
 - Sebastiano Lewental → 1 ora
- **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

- Presentazione di una demo del software;
- Confronto sul cambio di tecnologie;
- Range valori delle *metriche di prodotto*_o.

2.2 Argomenti e temi dell'incontro

2.2.1 Dimostrazione di una demo del software

Descrizione: I programmatori hanno mostrato a schermo una demo al fine di capire se le funzionalità fino ad ora sviluppate fossero state comprese nella loro intezza e, di conseguenza, fossero state sviluppate correttamente. Le funzionalità che sono state provate e mostrate inizialmente sono state:

- La possibilità di mostrare all'utente il *dizionario dati*_o in linguaggio naturale in modo tale che sia più comprensibile all'utente che interrogazioni può fare;
- L'inserimento di nuovi dizionari dati.

Risposta: E' stato data l'approvazione per la strutturazione generale delle funzionalità.

Descrizione: Sono state mostrate le funzionalità di generazione di un *prompt*_o successivo alla scelta di un *dizionario dati*_o e l'inserimento di una richiesta in linguaggio naturale il tutto in lingua inglese dato che il modello utilizzato fino ad ora utilizza questa lingua.

Risposta: La Proponente ha consigliato di utilizzare Bertino come *modello*_o in lingua italiana.

Descrizione: Il gruppo ha già lavorato con Bertino in precedenza ma i risultati degli score per la scelta delle tabelle per il *prompt*_o erano di molto superiori rispetto ad altri modelli e quindi sembrava che Bertino semplificasse di molto la scelta e i risultati sembravano essere troppo uniformati.

Risposta: La proponente ha quindi consigliato di usare Marco-BERTino dato che il capitolato esplicitava l'interesse per la lingua italiana. Viene spiegato che è importante avere la descrizione in italiano per non far ricadere la formulazione della query da parte degli *LLM*_o sulla descrizione delle tabelle restituite ovvero il loro lato semantico a causa di tutto il training che i modelli hanno attuato su frasi SQL in inglese.

Descrizione: E' stata testata una frase per la generazione del *prompt*_o.

Risposta: La proponente ha approvato la funzionalità cardine e ha proposto dei test da fare che sono stati provati durante il colloquio e da riprodurre in seguito. Inizialmente è stato proposto di usare il *prompt*_o generato dall'applicativo su chat.Im-sys.org per testare le *query*_o con diversi modelli e non solo con ChatGpt per vedere se il *prompt*_o fosse abbastanza chiaro anche per modelli LLM più piccoli.

Descrizione: Il *prompt*₆ generato è stato testato con 7 modelli di LLM diversi (chatGpt, yi-1.5-34b, qwen2-72b, claude-3-haiku, command-r, llama-3-8b, openhermes-mistral). I risultati emersi hanno confermato la correttezza del *prompt*₆ generato in quanto tutti i modelli hanno restituito la stessa *query*₆ in SQL se non per l'inserimento della maiuscola per i nomi propri che alcuni modelli hanno dato come preconditione anche se non era specificato nella frase inserita dall'utente né dal *prompt*₆ generato.

Risposta: Visti i buoni risultati la Proponente chiede di cambiare anche il DBMS e non usare solo MariaDB ma anche Postgres.

Descrizione: Dato il test di questa funzionalità si approfondisce il fatto che potrebbe essere una funzionalità aggiuntiva da implementare e da applicare come caso d'uso come la selezione della lingua in cui inserire la richiesta in linguaggio naturale.

Risposta: La Proponente risulta interessata a questa feature e ne appoggia l'implementazione.

Descrizione: Si testano altre *query*₆ e si analizza la struttura del *prompt*₆.

Risposta: La proponente evidenzia una problematica in quanto nel *prompt*₆ restituito fino ad ora vengono riportati solo i nomi delle tabelle e non le descrizioni ma in un caso realistico i nomi delle colonne nel *database*₆ non sono significativi e necessitano di una descrizione. Il *prompt*₆ restituito è quindi da modificare e ampliare.

Descrizione: Alla luce della nuova struttura del *prompt*₆ da restituire viene chiesto se sia possibile ottimizzare restituendo solo le descrizioni dei nomi di colonne non significativi e risparmiare quelli parlanti da sé.

Risposta: La Proponente rassicura che quest'ultima sarebbe una ottimizzazione non significativa in quanto per esempio ChatGpt gestisce 128K di richiesta e Gemini 1Mln quindi si possono restituire senza discriminante le descrizioni di tutte le tabelle. Sarebbe inoltre difficile decidere con quale criterio selezionare quali colonne descrivere e quali no.

Descrizione: Si passa a testare delle richieste in linguaggio naturale più articolate e meno specifiche.

Risposta: La Proponente mette in luce che sul *prompt*₆ è stata data più importanza a trovare la soluzione rispetto alla più precisa, è stata privilegiata la recall rispetto alla precisione. E' stata una scelta però necessaria che la Proponente si aspettava in quanto il problema si sarebbe presentato se avessimo scambiato l'ordine di queste due caratteristiche.

Descrizione: Viene mostrata la funzionalità di *debugging*₆ che è sviluppata su frasi di test di cui viene tratto il punteggio che hanno preso le tabelle e le colonne di quest'ultime per spiegare perché sono state selezionate certe parti di *database*₆ rispetto ad altre. Il debugger restituisce dei punteggi ,score, per risalire al meccanismo di selezione del *prompt*₆.

Risposta: Ciò che qui viene restituito è quello che si aspetta la Proponente di trovare anche nella generazione del *prompt*₆ ovvero le descrizioni di tabelle e colonne per il resto la funzionalità risulta essere corretta.

2.3 Cambio di Tecnologie

Descrizione: Terminata la dimostrazione della demo il gruppo ha condiviso di aver deciso di cambiare le tecnologie.

Risposta: La Proponente ha accolto l'informazione ribadendo che rimane indifferente alle scelte tecnologiche ma che la scelta sembra essere stata maturata dopo una valutazione profonda anche delle problematiche che si sarebbero potute presentare in futuro come era stato consigliato da quest'ultima negli incontri precedenti.

2.4 Range valori delle metriche di prodotto

Descrizione: Il gruppo ha voluto capire se, nei mesi precedenti durante l'affiancamento dei lavori dei gruppi che hanno preceduto Argo nello sviluppo di Chat SQL, le conclusioni a cui si è arrivati siano state abbastanza precise per capire come si potrebbero valutare le *metriche di qualità*, e che range selezionare per valutare il prodotto.

Risposta: La Proponente evidenzia come gli LLM evolvono a velocità esponenziali e ciò che mesi fa sembrava essere molto difficile ora probabilmente è già stato fatto. Settare dei range per la qualità del prodotto su questo capitolato ora sembra essere più semplice di prima in quanto è emerso come sia più semplice, e quindi fattibile, del previsto generare codice SQL corretto dato un *prompt*, di partenza. La Proponente invita quindi il gruppo, date queste premesse, a provare *query*, più complesse anche con subqueries nei mesi successivi per vedere se si possono raggiungere comunque buoni risultati come con quelle testate fino ad ora.

Descrizione: Vengono illustrati i test che si stanno implementando per la verifica del prodotto. La Proponente illustra un possibile test che può essere implementato in futuro e si testa una prima volta durante il colloquio.

Risposta: La Proponente chiede di prendere la frase in linguaggio naturale inserita dell'utente e farsela riscrivere da ChatGpt in modo differente e poi confrontare il prompt generato e la frase SQL che ne deriva. Se quest'ultime coincidono è un buon parametro di valutazione. La Proponente consiglia di ripetere il test con un modello in locale con 5 frasi diverse generate da una comune da un LLM.

3 Todo / In Progress

Ticket	Incarico	Incaricato/a	Scadenza
ARGO-130	Modificare dizionario dati [descrizioni tabelle e chiavi esterne]	Riccardo Cavalli	2024-06-09
ARGO-131	Aggiornare il prompt generator	Riccardo Cavalli	2024-06-09
ARGO-134	Test dizionario dati in italiano [modelli di efederici]	Riccardo Cavalli	2024-06-10
ARGO-136	Test con modelli locali	Riccardo Cavalli	2024-06-03

Luogo e Data:
Padova (PD) 2024-05-22

Firma: _____

Responsabile: Martina Dall'Amico

Per approvazione:

Firma: _____

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