



Basi di Dati (14AFQPL, 14AFQPI)
Anno Accademico 2024-2025
Politecnico di Torino

Testing di Large Language Model: Text2SQL – Create Table

Obiettivo

La finalità di questa esercitazione è quella di, dato un argomento, generare coppie domande-risposte relative al task Text2SQL – Create Table (ovvero dato un testo in linguaggio naturale formulare il corrispondente codice di Create Table) associabili a pattern di difficoltà differenti, e di testare le abilità di tre differenti Large Language Model (LLM), confrontando soluzioni generate con diversi prompt con la soluzione attesa.

Argomenti

L'argomento da utilizzare per la creazione del testo da usare nei prompt è da derivare dalla seguente tabella, divisa per turno di laboratorio:

TURNO	ARGOMENTO
Corso prof. Cagliari 11:30-13:00	Biblioteca
Corso prof. Cagliari 13:00-14:30	Università
Corso prof.ssa Farinetti 16:00-17:30	E-commerce
Corso prof.ssa Farinetti 17:30-19:00	Hotel

Individuato l'argomento assegnato, bisogna utilizzare un sottoinsieme delle entità elencate nel seguito e relazioni a piacere per scrivere i testi che rispettano i pattern richiesti

- **Argomento:** Biblioteca
 - Entità: Libro, Sala, Operatore, Lettore, Prestito, Evento, Autore, Evento, Ditta pulizia, Ospite, Prenotazione
- **Argomento:** Università
 - Entità: Studente, Docente, Corso, Operatore, Evento, Progetto, Lezione, Aula, Esame, Ticket, Dipartimento
- **Argomento:** E-commerce
 - Entità: Prodotto, Utente, Ordine, Spedizione, Fornitore, Corriere, Venditore, Negozio, Promozione, Recensione, Magazzino
- **Argomento:** Hotel
 - Entità: Hotel, Impiegato, Cliente, Stanza, Servizio, Prenotazione, Fornitore, Recensione, Congresso, Ristorante, PacchettoVacanza

Svolgimento

I form da compilare per l'esercitazione sono predisposti per memorizzare i seguenti campi:

- **Pattern di difficoltà:** pattern caratterizzante il testo proposto
- **Testo:** testo proposto in linguaggio naturale
- **Soluzione attesa:** soluzione attesa rispetto alla domanda proposta (codice Create Table)
- **Lingua:** lingua utilizzata per il prompt (italiano o inglese)
- **Prompt:** testo da inviare al LLM
- **Codestral/GPT-4o/Gemini:** campi nei quali inserire il codice SQL generato come risposta dai rispettivi LLM
- **Analisi Codestral/GPT-4o/Gemini:** commento dello studente alla soluzione proposta dai rispettivi LLM

Al fondo del secondo form è presente un campo nel quale lo studente può riportare un commento generale sulle principali risultanze sperimentali evidenziate tra cui, ad es.,

- Punti forti e punti deboli di ciascun LLM
- Confronti tra LLM
- Analisi e confronti tra prompt
- Confronto tra lingue

Durante il laboratorio a ciascun studente è richiesto di:

- Creare **un testo per ogni tipologia di pattern** indicata nel form.
- Scrivere la soluzione attesa al testo proposto.
- Formulare i prompt testuali da inviare all'LLM. Occorre partire dalla struttura dei prompt forniti in calce per ogni tipologia di pattern richiesto. Le caratteristiche del prompt dipendono **dalla lingua** (italiano o inglese) e **dal tipo**:
 - **Zero-Shot** Learning (ZSL): al modello non viene fornito nessun esempio
 - **Few-Shot** Learning (FSL): vengono forniti K esempi al modello. Per questo laboratorio K=3, cioè dovranno essere forniti **3 esempi**.

Nota: per alcuni pattern (0, 6, 7, 8) occorre generare sia prompt di tipo Zero-Shot che di tipo Few-Shot; per i restanti pattern basta generare i prompt di tipo Zero-Shot. Per tradurre i prompt dall'italiano all'inglese potete procedere manualmente oppure utilizzare un tool di traduzione automatica (DeepL o Google Traduttore)

- Interrogare **i tre modelli (LLM)** utilizzando i prompt generati:
 - **Codestral:** <https://chat.mistral.ai/chat>
 - **GPT-4o:** <https://chatgpt.com/>
 - **Gemini:** <https://gemini.google.com/app>
- Inserire l'output fornito dai vari LLM all'interno dei campi di testo delle corrispondenti domande del form **seguendo obbligatoriamente** gli spazi forniti (per l'esercizio 0, che bisogna svolgere, si può prendere spunto dagli esempi di prompt a fine testo).
- **Analizzare** le risposte fornite dai modelli e riportare i relativi commenti nei campi predisposti.

Per poter usare i vari modelli, occorre registrarsi tramite la propria e-mail personale (potete usare l'indirizzo studenti fornito da PoliTo) o effettuare l'accesso tramite account terzi (e.g. Google).

Lo svolgimento del laboratorio prevede **due fasi distinte** ciascuna delle quali consente di ottenere i punti del primo Homework. Il punteggio totale massimo assegnabile al primo Homework è **1 punto, diviso equamente tra le due fasi**.

Fase 1 - in laboratorio (0.5 punti)

- Seguendo le indicazioni dell'esercitatore di laboratorio, collegarsi alle interfacce Web di ciascun LLM e generare le coppie domande-risposte con i relativi prompt, soluzioni e commenti.
- Prima del termine dell'esercitazione, inviare il form adibito al laboratorio compilato arrivando all'ultima pagina e premento **INVIA**.
 - Se il form non è stato completato nello slot di laboratorio assegnato, **salvare** le risposte inviate cliccando sul pulsante "Salva la mia risposta per modificare" mostrato dopo l'invio del modulo. Se non si effettua il salvataggio delle risposte inviate, non sarà possibile continuare la compilazione in un secondo momento.
 - Per continuare la compilazione, accedere a Microsoft Forms da forms.office.com. Si troverà il modulo inviato nella scheda "Recente" o "Moduli compilati".
- La consegna è ritenuta valida **se e solo se**:
 - è stato inviato;
 - contiene i testi-risposte e i relativi prompt, soluzioni, e commenti degli **esercizi da 0 a 3**;
 - Data e orario di consegna non superano l'orario di conclusione del proprio slot di laboratorio.
- La **consegna del primo form** può essere effettuata in modo **individuale** o **a coppie**.
 - **Individuale**: compilare i campi "Nome/Cognome/Matricola/Email (1)", lasciando vuoti i campi "Nome/Cognome/Matricola/Email (2)".
 - **A coppie**: compilare sia i campi "Nome/Cognome/Matricola/Email (1)" che i campi "Nome/Cognome/Matricola/Email (2)".

Se non si completa il primo form durante lo slot di laboratorio, è necessario completarlo (sempre a coppie) a casa.

Fase 2 – completamento del lavoro (0.5 punti)

- **Completare il primo form qualora non già completato in laboratorio con la stessa modalità utilizzata in laboratorio** (se svolto in modo individuale continuare in modo individuale, se svolto a coppie continuare a coppie).
- Completare il secondo form adibito allo svolgimento a casa inserendo anche i **commenti finali**.
- La **consegna del secondo form** può essere effettuata esclusivamente in modo **individuale**.
- **Entro la scadenza definitiva** inviare il secondo form compilato.
 - Scadenza: 21 dicembre 2024 ore 23:59.
- La consegna è ritenuta valida **se e solo se**:
 - è stato inviato;
 - contiene i testi-rispo, le soluzioni, i prompt richiesti e i commenti finali;
 - la data di consegna non supera la scadenza definitiva.

Note

Lo svolgimento del quaderno con un topic diverso da quello assegnato comporta la mancata valutazione dell'elaborato. Tutte le sottomissioni vengono valutate, anche se parziali, con eventuali punteggi discretizzati.

Pattern

Le query proposte seguono livelli di difficoltà crescente, secondo i seguenti pattern:

0. [LAB] Due entità con relazione binaria uno a molti, con attributi anche nella relazione
1. [LAB] Due entità con relazione binaria molti a molti, con attributi anche nella relazione
2. [LAB] Due entità di cui una debole, e una relazione
3. [LAB] Generalizzazione (totale, esclusiva) con padre ed almeno 2 figli
4. [CASA] Generalizzazione (parziale, sovrapposta) con padre ed almeno 2 figli
5. [CASA] Entità con un attributo multi-valore
6. [CASA] Relazione ternaria col tempo (due entità più entità tempo)
7. [CASA] Relazione binaria col tempo (una entità più entità tempo)
8. [CASA] Entità storicizzata più altre due entità, con le relazioni necessarie
9. [CASA] 0 + 3 + 6 + 8
10. [CASA] 1 + 4 + 7 + 8

Piattaforma

Per svolgere il quaderno, dovete compilare i seguenti due form, accessibile esclusivamente attraverso la mail del Politecnico:

- **Primo form** (parte in **laboratorio**): <https://forms.office.com/e/Pm1WJjtqvH>
Scadenza: termine slot laboratorio
- **Secondo form** (parte **a casa**): <https://forms.office.com/e/qPXgJQ1KaQ>
Scadenza: 21 dicembre 2024 ore 23:59

Per poter inviare e sottomettere correttamente le risposte, occorre arrivare all'ultima pagina e premere il pulsante **INVIA**.

Sottomissioni precedenti all'orario di inizio del proprio slot o dopo la scadenza non verranno considerate.

Esempi di prompt

Seguono gli esempi sulla base del pattern 0 di esempio:

LLM prompting con Zero-Shot Learning:

“Sei un assistente AI per la progettazione di database partendo da una descrizione testuale.

Ti verrà fornita in input la descrizione dello scenario da rappresentare.

Tu dovrai fornire come output solamente il codice SQL per effettuare la creazione delle tabelle necessarie (CREATE TABLE).

Testo: Un allenatore può effettuare diverse sessioni di allenamento. Quando si effettua una sessione di allenamento, memorizzare l'informazione della durata.

Codice CREATE TABLE:

”

LLM Prompting con Few-Shot Learning:

“Sei un assistente AI per la progettazione di database partendo da una descrizione testuale.

Ti verrà fornita in input la descrizione dello scenario da rappresentare.

Tu dovrai fornire come output solamente il codice SQL per effettuare la creazione delle tabelle necessarie (CREATE TABLE).

Esempi:

Testo 1:

Un allenatore può effettuare diverse sessioni di allenamento. Quando si effettua una sessione di allenamento, memorizzare l'informazione della durata.

Codice CREATE TABLE 1:

```
CREATE TABLE Allenatori (  
    id_allenatore INT PRIMARY KEY,  
    nome VARCHAR(100) NOT NULL  
);
```

```
CREATE TABLE SessioniAllenamento (  
    id_sessione INT PRIMARY KEY,  
    id_allenatore INT,  
    durata INT NOT NULL, -- Durata in minuti  
    FOREIGN KEY (id_allenatore) REFERENCES Allenatori(id_allenatore)  
);
```

Testo 2:

Un giocatore gioca in una squadra. Quando si memorizza in che squadra gioca, salvare il costo d'ingaggio.

Codice CREATE TABLE 2:

```
CREATE TABLE Squadre (  
    id_squadra INT PRIMARY KEY,  
    nome_squadra VARCHAR(100) NOT NULL  
);  
CREATE TABLE Giocatori (  
    id_giocatore INT PRIMARY KEY,  
    nome_giocatore VARCHAR(100) NOT NULL,  
    id_squadra INT,  
    costo_ingaggio DECIMAL(10, 2),  
    FOREIGN KEY (id_squadra) REFERENCES Squadre(id_squadra)  
);
```

Testo 3: Una squadra vince un campionato. Memorizzare l'anno di vittoria del campionato.

Codice CREATE TABLE 3:

```
CREATE TABLE Squadre (  
    id_squadra INT PRIMARY KEY,  
    nome_squadra VARCHAR(100) NOT NULL  
);  
CREATE TABLE Campionato (  
    id_campionato INT PRIMARY KEY,  
    nome_campionato VARCHAR(100) NOT NULL,  
    id_squadra INT,  
    anno_campionato INT,  
    FOREIGN KEY (id_squadra) REFERENCES Squadre(id_squadra)  
);
```

Testo: Una squadra fallisce. Memorizzare l'anno del fallimento.

Codice CREATE TABLE:”