



# Code7Crusaders

Software Development Team

Glossario

## **Membri del Team:**

Enrico Cotti Cottini, Gabriele Di Pietro, Tommaso Diviesti  
Francesco Lapenna, Matthew Pan, Eddy Pinarello, Filippo Rizzolo

Versioni

Ver	Data	Redattore	Verificatore	descrizione
0.2	18/11/2024	Enrico Cotti Cottini		Aggiunta di nuovi termini
0.1	05/11/2024	Gabriele Di Pietro	Filippo Rizzolo	Prima stesura del documento

## Indice

## A

### **Analisi dei Requisiti**

Processo di identificazione e definizione delle necessità e delle aspettative degli stakeholder per un progetto. Serve come base per la progettazione e lo sviluppo del prodotto.

### **API REST (Representational State Transfer)**

Un'architettura per la progettazione di servizi web che consente la comunicazione tra sistemi attraverso il protocollo HTTP. In un'API REST, le risorse (come dati o funzionalità) sono identificate da URL univoci e possono essere manipolate utilizzando operazioni standard come GET, POST, PUT e DELETE. L'architettura è stateless, il che significa che ogni richiesta contiene tutte le informazioni necessarie per essere elaborata senza fare affidamento su dati memorizzati sul server. È ampiamente utilizzata per creare applicazioni scalabili e interoperabili.

## B

### Backend

La componente di un'applicazione software responsabile della logica di business, della gestione dei dati e dell'elaborazione delle richieste provenienti dal frontend. Si occupa di eseguire operazioni sul database, elaborare algoritmi complessi e garantire il corretto funzionamento delle funzionalità. Comunica con il frontend principalmente tramite API, fornendo i dati necessari per l'interfaccia utente.

Un benchmark è un punto di riferimento utilizzato per misurare le prestazioni di un sistema, di un'applicazione, o di un dispositivo rispetto a un set di standard predefiniti o a sistemi comparabili. Nella tecnologia, un benchmark può riferirsi a test eseguiti per valutare l'efficienza, la velocità, la stabilità, e altre caratteristiche importanti di hardware, software, o processi aziendali. Questi test sono cruciali per identificare aree di miglioramento, confrontare soluzioni concorrenti, e garantire che un sistema funzioni in modo ottimale rispetto alle aspettative del mercato o alle specifiche di progetto.

### BERT (Bidirectional Encoder Representations from Transformers)

BERT è un modello di linguaggio sviluppato da Google che utilizza trasformatori per comprendere il contesto di una parola considerando sia le parole precedenti che quelle successive. Questo approccio bidirezionale migliora la comprensione del testo in compiti come la classificazione, la risposta a domande e l'analisi del linguaggio naturale. BERT è pre-addestrato su grandi dataset e può essere facilmente adattato a diverse applicazioni specifiche.

## D

### Database Relazionale

Un sistema di gestione dei dati che organizza le informazioni in tabelle strutturate composte da righe e colonne. Le tabelle sono correlate tra loro attraverso chiavi primarie e chiavi esterne, consentendo di mantenere l'integrità dei dati e facilitare l'accesso e la manipolazione. Utilizza il linguaggio SQL (Structured Query Language) per eseguire operazioni come la creazione, l'interrogazione, l'aggiornamento e l'eliminazione dei dati. I database relazionali sono ampiamente utilizzati in applicazioni aziendali e web per la loro efficienza e scalabilità.

### Database Vettoriale

Un database vettoriale è un tipo di sistema di gestione dei dati progettato per memorizzare, indicizzare e cercare vettori, ossia rappresentazioni numeriche di oggetti come parole, immagini o concetti. I vettori sono utilizzati per catturare le caratteristiche semantiche di un dato e vengono comunemente usati nei modelli di machine learning e di linguaggio naturale. I database vettoriali permettono di effettuare ricerche rapide basate sulla similarità tra vettori, come nel caso della ricerca per vicinanza (nearest neighbor search), che è utile per applicazioni come il recupero delle informazioni e la raccomandazione di contenuti.

### Diagrammi UML

Diagrammi standardizzati utilizzati per modellare e visualizzare il design di sistemi software. Aiutano a descrivere la struttura, il comportamento e le interazioni tra componenti del sistema.

### Documentazione Formale e Informale

La documentazione formale include materiali ufficiali come norme e piani di progetto, mentre quella informale comprende appunti e bozze non ancora approvate. Entrambe supportano lo sviluppo del progetto.

## E

### Embedding

Una tecnica utilizzata nell'intelligenza artificiale e nell'elaborazione del linguaggio naturale per rappresentare parole, frasi o interi documenti in uno spazio vettoriale. Ogni elemento viene trasformato in un vettore numerico, dove la posizione e la distanza tra i vettori riflettono il significato semantico. Gli embedding consentono di confrontare testi in modo efficiente, trovando similarità tra concetti e facilitando attività come la ricerca semantica, la classificazione dei testi e il clustering.

## **F**

### **FAISS**

Una libreria open-source sviluppata da Facebook AI Research per l'indicizzazione, la ricerca e il recupero efficiente di vettori in spazi di grandi dimensioni. FAISS (Facebook AI Similarity Search) è progettata per gestire nearest neighbor search in dataset di grandi dimensioni, garantendo alta velocità ed efficienza. Viene utilizzata in applicazioni come la ricerca semantica, i sistemi di raccomandazione e l'elaborazione del linguaggio naturale.

### **Frontend**

La parte visibile di un'applicazione software che rappresenta l'interfaccia utente (UI) e consente agli utenti di interagire con il sistema. È responsabile della presentazione grafica, della navigazione e dell'acquisizione delle informazioni inserite dagli utenti. Il frontend comunica con il backend attraverso API per inviare e ricevere dati, garantendo un'esperienza utente fluida e interattiva.



## G

### GitHub Actions

Una piattaforma di automazione che permette di eseguire flussi di lavoro direttamente nei repository GitHub. Supporta azioni come compilazioni, test e deployment automatizzati, migliorando l'efficienza nello sviluppo software.

### Glossario

Elenco strutturato di termini tecnici o specializzati, ognuno corredato dalla propria definizione o spiegazione. Questo strumento aiuta a migliorare la comunicazione tra le varie parti coinvolte in un progetto, riducendo le ambiguità e garantendo una comprensione condivisa dei termini utilizzati in un determinato contesto.

## **L**

### **LangChain**

LangChain è una libreria open-source progettata per semplificare l'integrazione e l'orchestrazione di modelli di linguaggio avanzati (come GPT) con altre fonti di dati e sistemi. Viene utilizzata per costruire applicazioni complesse che combinano intelligenza artificiale e database, rendendo possibile la gestione di flussi di lavoro dinamici. LangChain permette di creare pipeline che automatizzano il processamento delle informazioni, la comunicazione con le API, l'elaborazione di documenti e l'interazione con vari strumenti esterni. Con un focus sull'accessibilità e la modularità, LangChain rende più semplice l'implementazione di soluzioni di AI che richiedono un'integrazione fluida tra diversi moduli e fonti di dati.

### **LLM (Large Language Model)**

Un modello di intelligenza artificiale addestrato su grandi quantità di dati testuali per comprendere, generare e contestualizzare il linguaggio naturale. È utilizzato in applicazioni come chatbot, traduttori e analisi del linguaggio.

# M

## Machine Learning

Una branca dell'intelligenza artificiale che utilizza algoritmi per apprendere dai dati e migliorare le prestazioni senza essere esplicitamente programmata. Trova applicazione in ambiti come la previsione, il riconoscimento e la personalizzazione.

## Milestone

Un obiettivo intermedio significativo nel ciclo di vita del progetto, utilizzato per monitorare i progressi e verificare il completamento di specifiche fasi o attività.

## N

### **Natural Language Processing (NLP)**

Tecnologia che combina linguistica e intelligenza artificiale per consentire ai computer di comprendere, analizzare e generare linguaggio naturale. Viene usata in chatbot, traduttori automatici e analisi testuale.

### **Nearest Neighbor Search (NNS)**

Un algoritmo utilizzato per trovare i punti più vicini in uno spazio vettoriale, spesso utilizzato in database vettoriali per recuperare embedding simili.

### **Norme di Progetto**

Regole e linee guida stabilite all'interno di un progetto per garantire coerenza e qualità nelle attività svolte. Definiscono standard e procedure, come documentazione, gestione delle versioni e criteri di codifica, per assicurare uniformità nell'approccio e nel risultato finale.

## **P**

### **PB (Product Baseline)**

Una revisione che documenta lo stato del progetto al termine dello sviluppo, descrivendo il prodotto finale in termini di specifiche tecniche, funzionalità e conformità ai requisiti iniziali.

### **Piano di Progetto**

Documento formale che delinea in dettaglio la pianificazione, la esecuzione, il monitoraggio e il controllo di tutte le attività coinvolte nella realizzazione di un progetto. Questo documento fornisce una roadmap chiara e organizzata, comprensiva di obiettivi, risorse, scadenze e strategie di gestione dei rischi.

### **Piano di Qualifica**

Documento che stabilisce gli standard di qualità, i processi e le attività di testing che saranno implementati durante lo sviluppo di un progetto. Contiene una descrizione dettagliata delle strategie di testing, delle metriche di valutazione e dei criteri di accettazione del prodotto finale.

### **Pipeline RAG**

Retrieval-Augmented Generation, un'architettura che combina il recupero di informazioni da un database con modelli di linguaggio per generare risposte accurate e contestualizzate.

### **PoC (Proof of Concept)**

Una dimostrazione pratica per validare la fattibilità di un'idea, tecnologia o progetto. Permette di identificare eventuali rischi o limitazioni prima di procedere con lo sviluppo completo.

### **PostgreSQL**

PostgreSQL è un sistema di gestione di database relazionali open-source, noto per la sua robustezza, scalabilità e supporto per operazioni avanzate come transazioni ACID e gestione di dati JSON. È ampiamente utilizzato per applicazioni che richiedono affidabilità e performance.

## Q

### Query Semantica

La query semantica è una tecnica di ricerca che si basa sull'analisi del significato delle parole nel linguaggio naturale, piuttosto che sulla semplice corrispondenza di termini. Permette di trovare informazioni rilevanti in un database tenendo conto del contesto e delle relazioni tra concetti, migliorando la qualità dei risultati. Viene utilizzata in motori di ricerca e assistenti virtuali per comprendere meglio le intenzioni dell'utente.

## **R**

### **Roadmap di GitHub**

Strumento visuale per pianificare e monitorare il progresso delle attività di un progetto all'interno di GitHub. Aiuta a gestire priorità, scadenze e assegnazione dei compiti.

### **RTB (Requirements and Technology Baseline)**

Una revisione formale che stabilisce i requisiti funzionali e tecnici di un progetto, definendo il punto di riferimento per lo sviluppo e la verifica delle funzionalità richieste.

## S

### Scalabilità

La scalabilità è la capacità di un sistema di gestire un aumento del carico di lavoro o delle risorse senza compromettere le prestazioni. Un sistema scalabile può adattarsi a una maggiore domanda aggiungendo risorse in modo efficace, sia aumentando la capacità di singole risorse (scalabilità verticale) che aggiungendo più risorse (scalabilità orizzontale). È essenziale per mantenere alte le performance in ambienti in crescita.



## **T**

### **Template di LaTeX**

Un modello predefinito per la creazione di documenti in LaTeX. Consente di standardizzare la struttura e il formato dei documenti, semplificando la redazione e migliorando la coerenza stilistica.

### **Token**

Un token è l'unità minima di testo elaborata da un modello di linguaggio. Può essere una parola intera, un simbolo, una parte di una parola o una sequenza di caratteri che il modello interpreta come un singolo elemento. I modelli di linguaggio suddividono il testo in token per analizzare e comprendere il contenuto, facilitando l'elaborazione del linguaggio naturale. La suddivisione in token è un passaggio cruciale nella comprensione e generazione del testo da parte dei modelli di intelligenza artificiale.

## V

### **Versionamento**

Processo di gestione delle modifiche ai documenti o al codice sorgente mediante strumenti che tracciano e memorizzano le varie versioni. Garantisce trasparenza e facilita la collaborazione tra membri del team.