Costruzione di un database sugli eventi attraverso l'utilizzo di un webcrawler e LLMs

Tesi di Agostino Vigani

Relatore: Andrea Maurino

Co-relatore: Blerina Spahiu

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Introduzione

Ipotesi di partenza

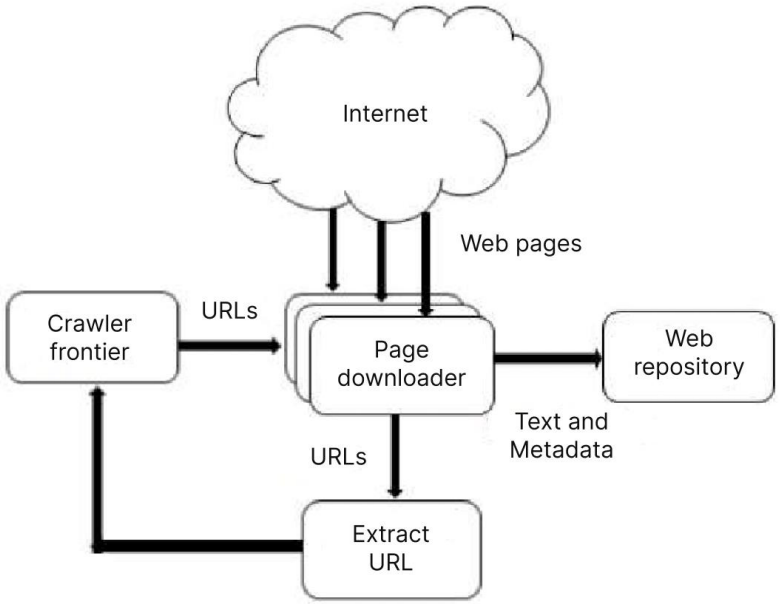
- Risolvere il problema della **frammentazione** delle informazioni online sugli eventi
- Offrire **dati** accurati per **applicazioni future**

Obiettivi

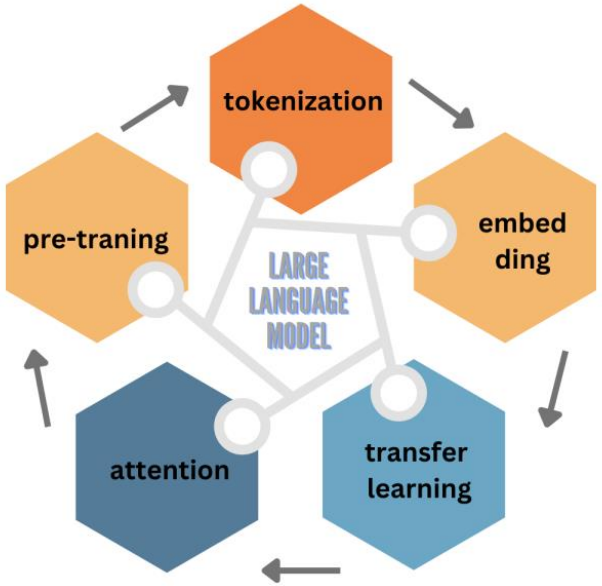
- Implementare un **sistema** efficiente per la raccolta dati
- Utilizzare tecniche di **web crawling e LLM**
- Valutare la **qualità** dei dati ottenuti

Tecnologie abilitanti

Web Crawler

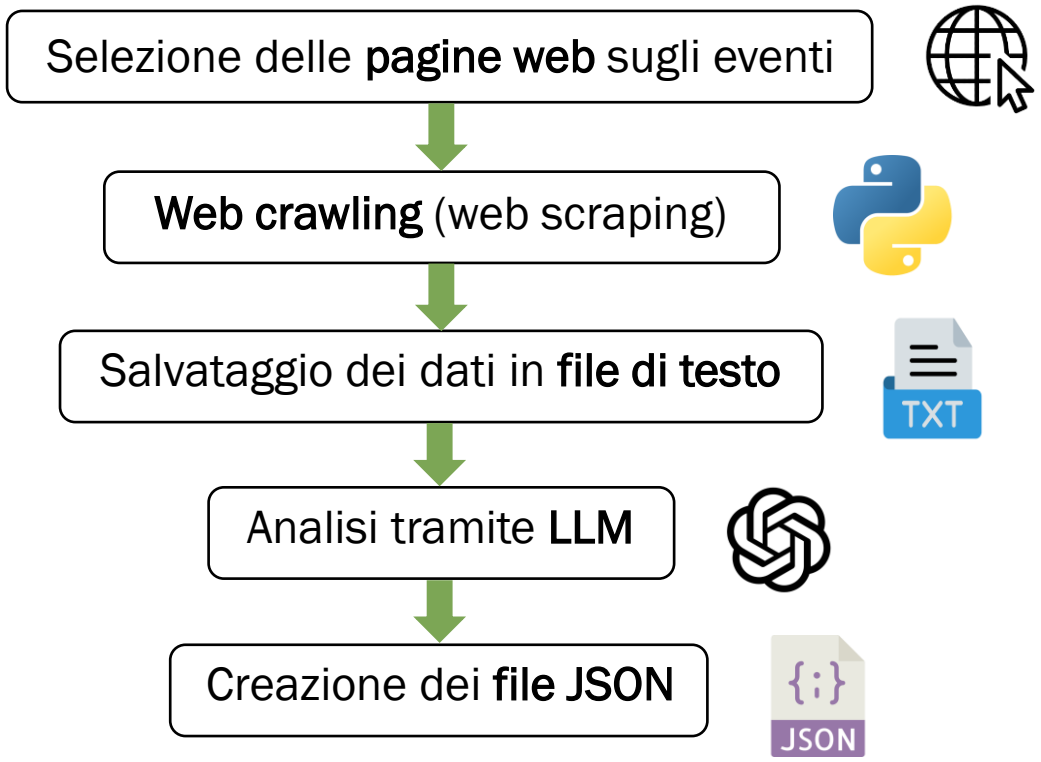


Large Language Model



UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Pipeline di sviluppo



Esperimenti


```
{
    "events": [
        {
            "title": "Ghemon",
            "url": "https://www.rockol.it/concerti-ghemon-a-wod5zzqbd54",
            "location": "Teatro degli Arcimboldi",
            "address": "Viale dell'Innovazione, 20, 20126 Milano MI, Italy",
            "date": [
                {
                    "date_day": "2024-11-08",
                    "date_hour": "21:00"
                }
            ],
            "category": "Concert"
        },
        {
            "title": "Mike Stern",
            "url": "https://www.rockol.it/concerti-mike-stern-a-79zdg4n4d0b",
            "location": "Blue Note",
            "address": "Via Pietro Borsieri, 37, 20159 Milano MI, Italy",
            "date": [
                {
                    "date_day": "2024-11-08",
                    "date_hour": "20:30"
                },
                {
                    "date_day": "2024-11-08",
                    "date_hour": "23:00"
                }
            ],
            "category": "Concert"
        },
        {
            "title": "Franco D'Andrea",
```

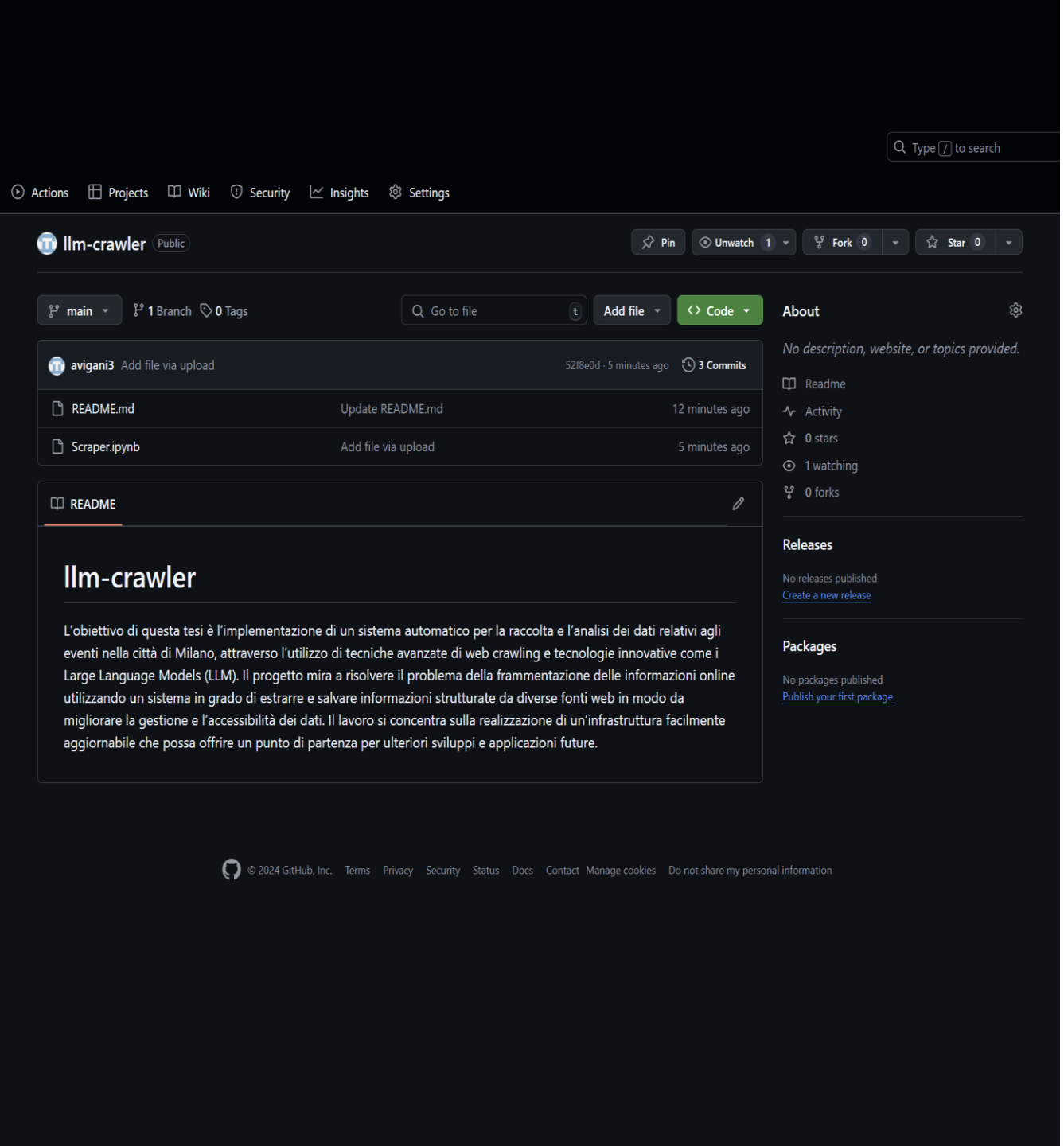

Valutazione delle metriche

- Conformità
- Completezza
- Accuratezza

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Risultati

Sito Web	N. Eventi	Conformità	Completezza	Accuratezza
Teatro alla Scala	142	Positivo	Positivo	5/5
Piccolo Teatro di Milano	53	Positivo	Positivo	5/5
Teatro Elfo Puccini	60	Positivo	Positivo	4/4
Teatro Litta	38	Positivo	Positivo	0/5
MUDEC - Museo delle Culture	4	Positivo	Positivo	3/3
Pinacoteca di Brera	10	Positivo	Positivo	3/3
Pirelli HangarBicocca	6	Positivo	Positivo	3/3
Autodromo Nazionale di Monza	5	Positivo	Positivo	3/3
Concerti a Milano	719	Positivo	Positivo	5/5

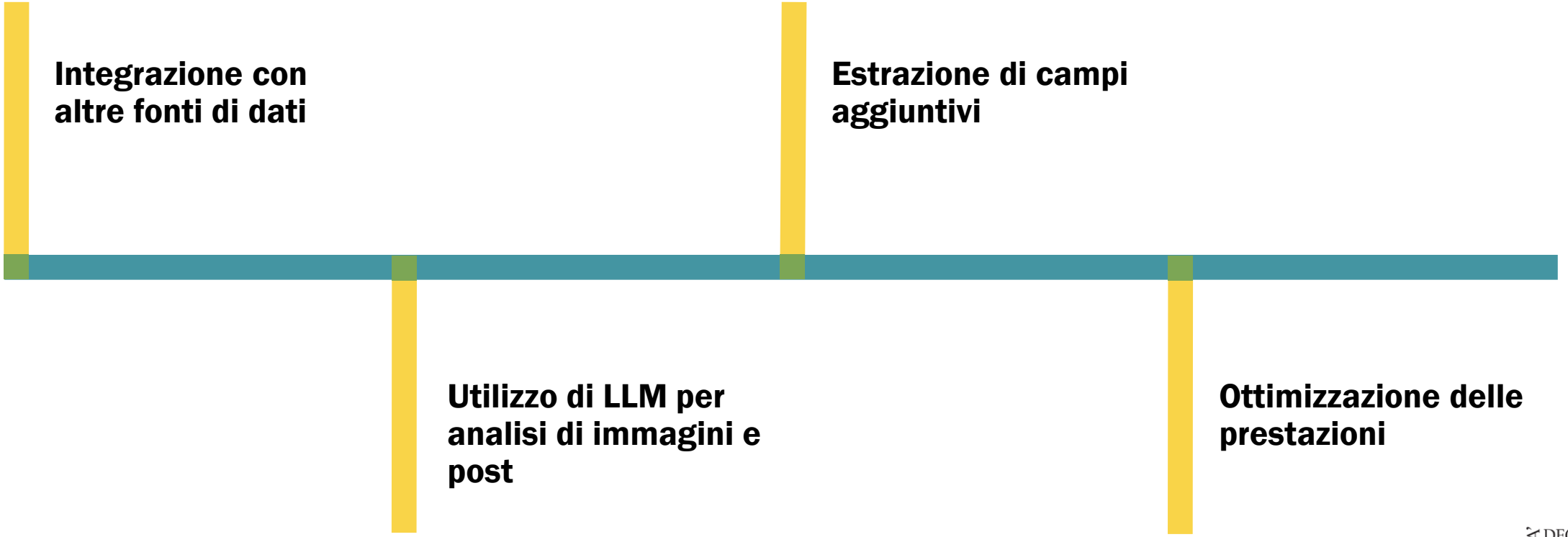


Conclusioni

- Implementazione di un **sistema** in grado di raccogliere **dati aggiornati** sugli eventi di Milano
- Utilizzo di un **approccio innovativo** insieme a tecnologie tradizionali per ottimizzare il lavoro
- Acquisizione di **competenze specifiche** nella programmazione, gestione dei dati e integrazione di modelli di AI

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Sviluppi futuri



UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA