

BriteUp

Search Engine di eventi

progetto per il corso di Gestione dell'Informazione dell'università di Modena e Reggio
Emilia anno 2021/2022

Daniele Bianchini e Nicholas Lo Piccolo



Indice

1. Idea
2. Sorgenti
3. Raccolta Documenti
4. Preprocessing
5. Schema degli indici
6. Linguaggio di querying
7. Search engine
8. Esempi di query di ricerca
9. Benchmark
10. Fine



Descrizione dell'idea

Il corso di gestione dell'informazione ci ha insegnato che i dati testuali sono fondamentali e hanno un potenziale unico.

Tutti i giorni ci imbattiamo in siti e search engine che sfruttano questi dati per mostrarci ricerche basate sui nostri interessi ottimizzando inserzioni e permettendoci di ridurre il tempo necessario per trovare quello che ci interessa.

L'idea alla base di BriteUp è quella di permettere ai suoi utenti di ricercare degli eventi di tutti i tipi.



Sorgenti

Le sorgenti sono due famosi siti per la creazione e la sponsorizzazione di eventi.

Eventbrite è un sito Americano che permette di gestire e vendere biglietti di eventi. Il servizio permette agli utenti di ricercare, creare e promuovere degli eventi.



Meetup invece punta a creare e promuovere gruppi sociali. permette ai gruppi di ritrovarsi tramite la creazione di eventi aperti a tutti gli interessati. Il sito permette di ricercare e creare eventi.

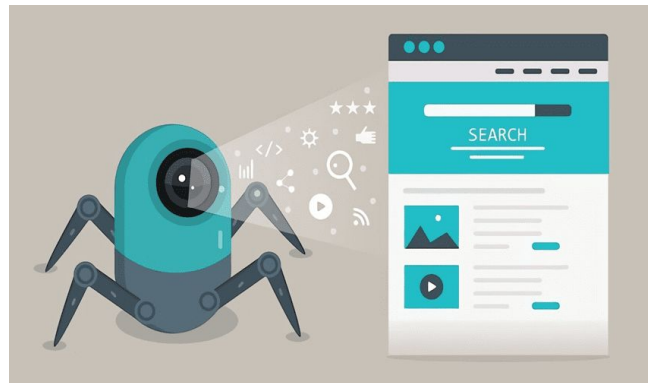
Raccolta dei documenti

Abbiamo sviluppato un **WebCrawler** in grado di indagare il sito andando a cercare link di tipo evento per tale motivo è stata sviluppata una funzione ad hoc per ogni sito sorgente in modo da poter fare scraping delle pagine html e creare i documenti con i dati utili ai fini di ricerca.

Nelle prime 3 righe di ogni documento possiamo trovare:

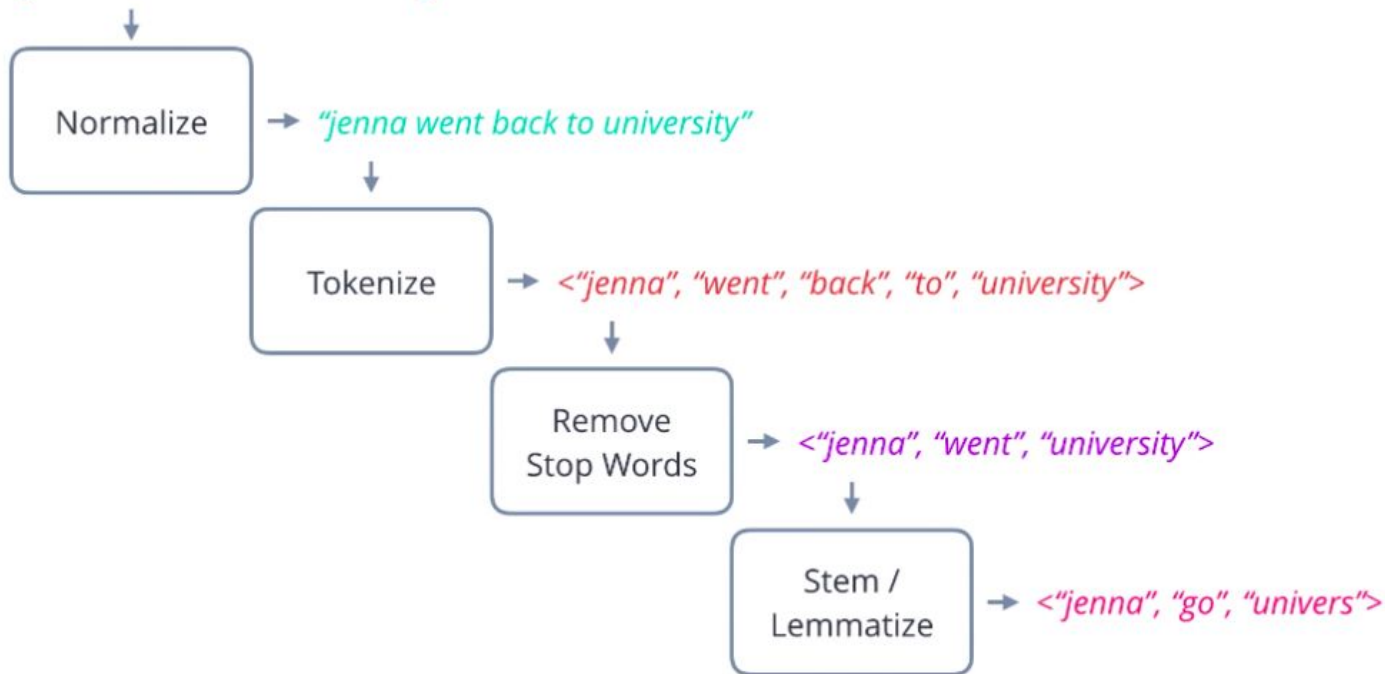
- Il link che ci riporterà al sito per avere una visione chiara dell'evento
- Il nome dell'evento
- Data e ora in cui si terrà l'evento

A seguire la descrizione dell'evento stesso.



PRE-PROCESSING

"Jenna went back to University."





Schema degli indici

Titolo Evento (event)	Descrizione (content)	Sito Web (url)	PATH
----------------------------------	----------------------------------	---------------------------	-------------



Linguaggio di querying

- Ricerca per keyword:

- movie

- country music

- Ricerca per field dello schema

- event: cosplay

- Composizione con operatori AND, OR

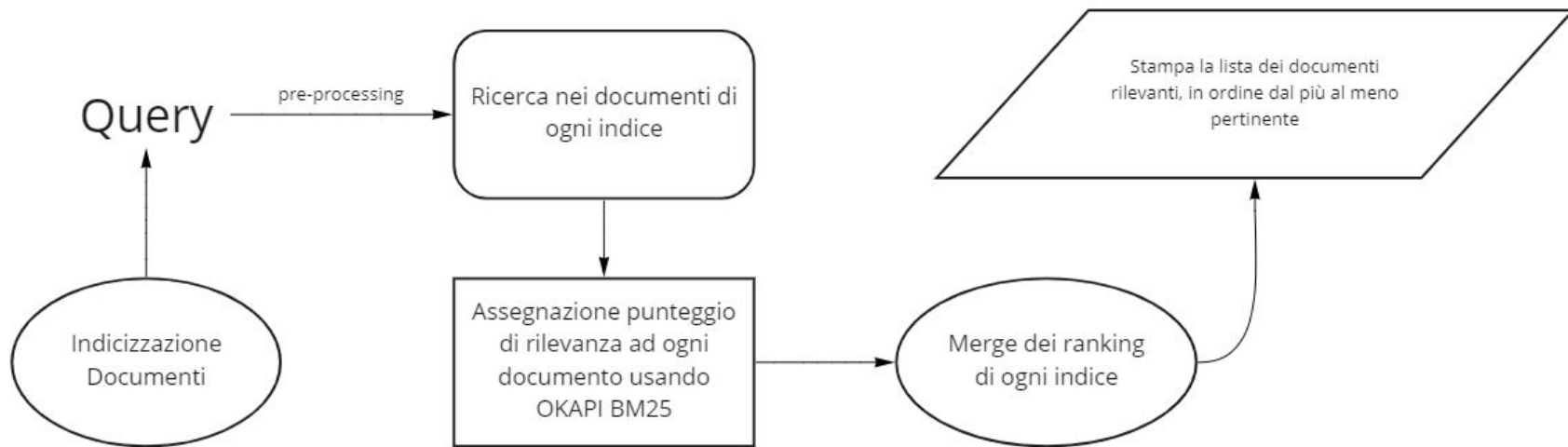
- music OR dance

- Query notevoli:

- ricerca di mese o giorno della settimana (tramite sigle) -> event:jan, event:mon



Search Engine





Esempi di query di ricerca

Le query di ricerca utilizzate sono le seguenti:

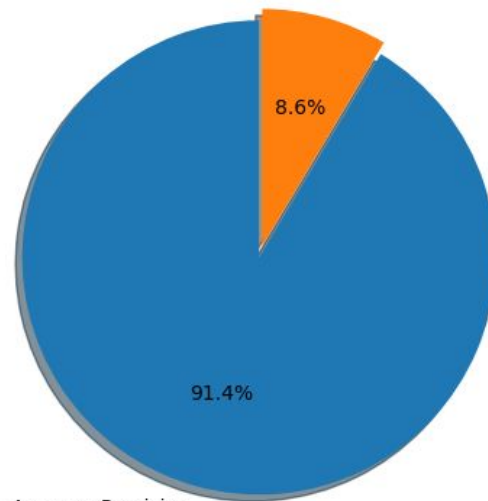
- karaoke
- dance
- speed dating
- science OR tech
- new york AND music
- content:comedy
- dating AND url:eventbrite
- london AND event:jun
- event:crypto AND event:blockchain
- event:jul AND (event:thu OR event:wed)



Benchmark

I seguenti risultato sono stati ottenuti in base alle query della slide precedente:

- 1
- 0.948
- 1
- 0.975
- 0.218
- 1
- 1
- 1
- 1
- 1



Mean Average Precision

Fine

Grazie per l'attenzione