



Progetto di Gestione dell'Informazione

Michele Bianco
Giovanni Anniballi

Introduzione

Il progetto che andremo a presentare è un search engine che ha come tema quello della cucina e delle ricette, un utente potrà cercare ricette filtrando i contenuti per titolo, ingredienti o categorie.

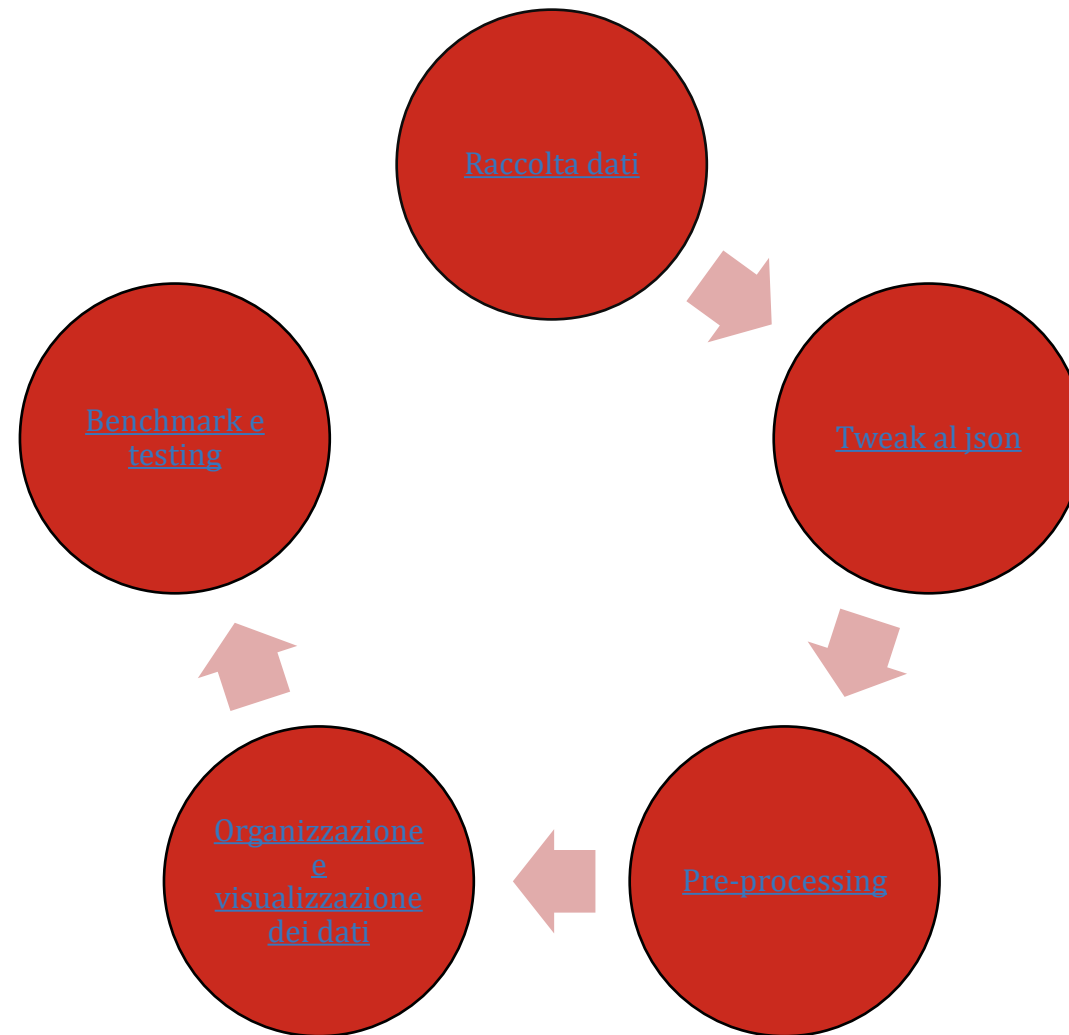
La differenziazione sulla ricerca è stata fatta per garantire la massima flessibilità all'utente, per esempio se si dispone solo di determinati ingredienti si potranno cercare ricette che contengono solo quei determinati ingredienti.

Premendo il tasto Sinonimi invece di Cerca sarà possibile appunto cercare i sinonimi della parola scritta nella barra di ricerca.

A dark-themed search bar interface. On the left, there is a dropdown menu with the text "Tutte" and a downward arrow. To its right is a large white rectangular input field containing the word "Cerca". Further right are two yellow rectangular buttons: the first is labeled "Cerca" and the second is labeled "Sinonimi".

Tutte ▼ Cerca Cerca Sinonimi

Fasi del lavoro



Raccolta dati

I dati sono stati raccolti da [giallozafferano](#) e da [donnamoderna](#) tramite un add-on chiamato Open Web Scraper (la cui documentazione può essere trovata [qui](#)).

Prima di iniziare la raccolta dati vera e propria abbiamo studiato cosa recuperare, capendo quanto i dati raccolti da due siti diversi potessero essere compatibili, per questo abbiamo deciso di selezionare i campi su cui fare scraping, i campi scelti sono stati:

- Fonte: contenente il link del sito dal quale proviene la ricetta
- Categoria: tipo di portata della ricetta (Primi Piatti, Contorno, Antipasti, etc.)
- Titolo: il nome della ricetta
- Ricetta-link: link alla ricetta che riporta alla pagina della fonte
- Ingredienti: comprensivi di grammatura di ogni prodotto
- Ingredienti_token: salva i prodotti in modo "pulito" in modo da favorire la ricerca su questo campo
- Preparazione: istruzioni da seguire per riprodurre la ricetta
- Immagine-src: una foto del piatto

Tweak del file json

Dopo lo scraping alcuni campi del json avevano bisogno di essere formattati per poter essere ritenuti leggibili, la funzione `json_fix()` serve a trasformare questo...

[illegible]

Tweak del file json

...in questo

```
{
  "fonte": "https://ricette.donnamoderna.com/",
  "categoria": "Primi piatti",
  "titolo": "Zuppa piccante al pomodoro",
  "ricetta-link": "https://ricette.donnamoderna.com/zuppa-piccante-al-pomodoro",
  "ingredienti_token": "cipolla, sedano, patata, carota, pomodoro, sale, peperoncino piccante, olio di oliva extravergine, origano, basilico",
  "ingredienti": "200 grammi cipolla, 200 grammi sedano, 400 grammi patata, 200 grammi carota, 1.5 chilogrammi pomodoro, q.b. sale, q.b. peperoncino piccante,
  \"preparazione\": \"1) Pulisci e taglia a tocchetti le patate e le verdure, tranne il peperone, mettetile in una casseruola, unisci mezzo litro di acqua, i pom
  \"immagine-src\": \"https://www.donnamoderna.com/content/uploads/2001/12/zuppa-piccante-al-pomodoro.jpg\"
},
```

Come si può osservare i campi cambiati sono ingredienti, ingredienti_token e preparazione.

Pre-processing e indexing

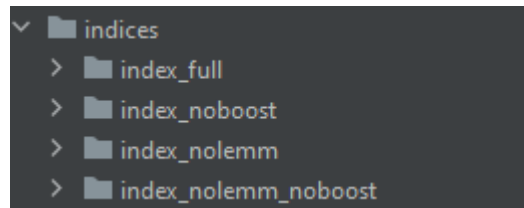
E' seguita poi la fase di pre-processing ed indicizzazione.

Per il pre-processing è stata usata la libreria NLTK tokenizzando il testo ed eliminando stopwords e punteggiatura, per la lemmatizzazione è stato utilizzato simplemma.

Successivamente è stata effettuata l'indicizzazione sulla base di uno schema con tre campi testuali principali utilizzati ai fini della ricerca: titolo, categoria e ingredienti_token.

Gli altri campi sono stati gestiti come "stored".

Sono stati creati diversi indici partendo da un indice completo (boost e lemmatizzazione) andando poi a sottrarre i singoli componenti del pre-processing. La differenza tra i vari indici creati sarà vista nei benchmark che mostreranno quale configurazione è la più efficiente.




Organizzazione e visualizzazione dei dati

Per mostrare i dati all'utente è stato utilizzato un modal che cerca di imitare un tradizionale ricettario cartaceo in modo da rendere più naturale l'utilizzo del sito anche ai più nostalgici.

Nella scheda possiamo trovare molti degli elementi salvati nel json, per esempio possiamo vedere immagine-src o ricetta-link, utilizzato nel tasto “Vai al sito” che reindirizza l'utente alla pagina della ricetta.

Pasta alla paolina



Ingredienti:

- Bucatini 320 g
- Acciughe sott'olio 20 g
- Cannella in polvere 1 cucchiaino
- Chiodi di garofano 3
- Aglione 1 spicchio
- Polpa di pomodoro 300 g
- Sale fino 1 cucchiaino
- Olio extravergine d'oliva q.b.
- Basilico q.b.
- Pangrattato 100 g

Preparazione

Per realizzare la pasta alla paolina prima cosa su un tagliere riducete in polvere i chiodi di garofano con l'aiuto di un batticarne 1 2 e tenete da parte. In un pentolino scaldate un giro di olio 3. Versate il pangrattato 4 e tostatelo fino a doratura 5. In un'altra padella scaldate un filo di olio extra vergine d'oliva, poi aggiungete lo spicchio di aglio intero 6 e fatelo rosolare. Eliminate l'aglio 7 e aggiungete le acciughe scolate 8. Scaldate a fuoco moderato mescolando, fino a quando non si saranno sciolte 9. Versate la polpa di pomodoro 10, aggiungete un cucchiaino di sale 11 e i chiodi di garofano tritati 12. Insaporite anche con un cucchiaino raso di cannella in polvere 13. Coprite con il coperchio 14 e lasciate cuocere a fiamma bassa per circa 10 minuti. Intanto tritate le foglie di basilico 15. Ponete sul fuoco un tegame colmo di acqua di cottura salata e quando raggiungerà il bollore versate gli spaghetti 16 e cuocetela al dente. Quando il sugo sarà pronto, insaporite con il basilico 17. Scolate poi la pasta nel tegame con il sugo 18. Mescolate e fate cuocere ancora un paio di minuti a fiamma bassissima 19. Versate la pasta in un piatto e spolverizzate con il pangrattato tostato 20. Completate la pasta alla Paolina con un pizzico di basilico tritato fresco, poi servitela ben calda 21.

[Vai al sito](#)

Linguaggio di interrogazione

E' stato utilizzato il linguaggio di interrogazione di Whoosh, che supporta operatori booleani e wildcards, nei benchmark abbiamo cercato di variare il più possibile nell'utilizzo di queste caratteristiche.

Alcuni esempi di nostri benchmark:

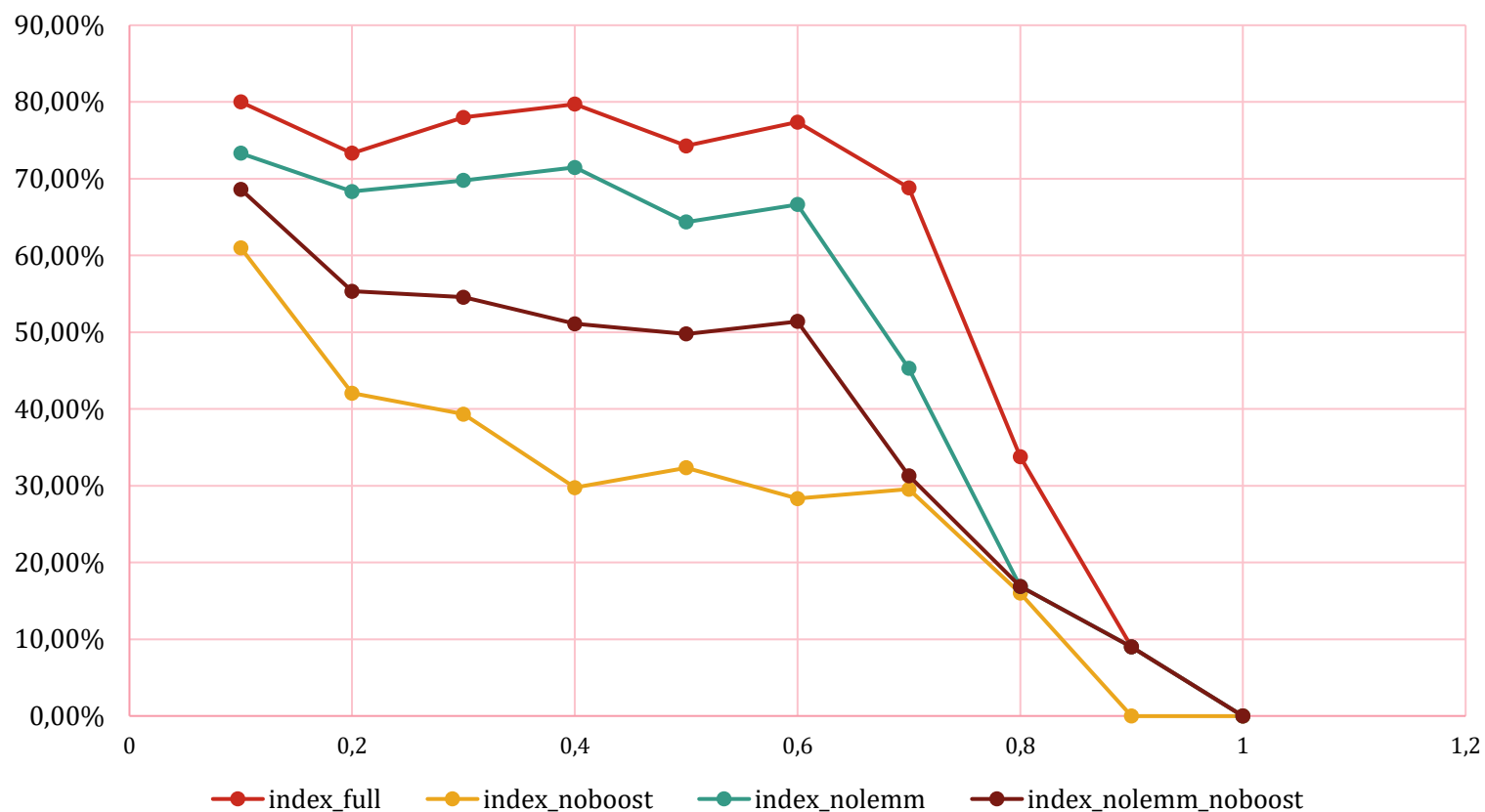
- categoria:pane OR pane e pizza
- ingredienti:panna
- pizza OR pane NOT pomodoro

La ricerca per titolo, categoria ed ingredienti è resa più semplice da un menù a tendina posto a sx della barra di ricerca. Per estendere le possibilità di query vi è un tasto apposito per la ricerca di sinonimi (tasto "Sinonimi").

La query immessa dall'utente, dopo un controllo di integrità (query non vuota o nulla), viene sottoposta a post-processing: le espressioni regolari precedute dal # verranno gestite come ricerca per sinonimi.

Benchmark

Come riferimento abbiamo deciso di utilizzare il motore di Ricerca **Google** in quanto è uno dei più famosi ed utilizzati per la ricerca di ricette.



Ranking

