Relazione

Palermo Cultural Guide

A cura di:

Buffa Salvatore 0656396

Curreri Dario 0651257

Gristina Salvatore Antonino 0652307

Indice

1. Introduzione TOTO’
2. Dati utilizzati e licenze IO
3. Definizione ontologia CURRERI
4. Papeline di elaborazione
   1. Unione dei dataset TOTO’
   2. Unione dataset con flickr CURRERI
   3. Interlinking TOTO’
   4. Creazione file a 5 stelle CURRERI / IO
   5. Bot telegram IO
5. Introduzione

IT:

“

Palermo Cultural Guide nasce con lo scopo di aiutare turisti e non solo, nella scoperta dei siti culturali di Palermo e provincia.

Lasciati guidare dal nostro bot, è necessario solamente inviare la tua posizione per ricevere descrizioni, localizzazione ed immagini dei siti culturali vicini a te! “

EN:

“

The goal of Palermo Cultural Guide is to help tourists and others, to discover the cultural sites

of Palermo and nearby.

Let our bot help you, you just need send your position to receive descriptions, locations and pictures of cultural sites near you! “

2. Dati utilizzati e licenze

Per realizzare questo progetto, sono stati utilizzati i seguenti dataset con relative licenze:

<http://dati.beniculturali.it/dataset/dataset-luoghiSicilia.json>

Licenza CC-BY 3.00:

Licenza di tipo Creative Commons che permette alll’usufruitore di distribuire, modificare e sviluppare anche commercialmente l'opera, riconoscendo sempre l'autore originale.

Licenza CC BY 4.0 IT:

Licenza di tipo Creative Commons che permette alll’usufruitore di distribuire, modificare e sviluppare anche commercialmente l'opera, riconoscendo sempre l'autore originale, presupponendo altresì l’attribuzione automatica di tale licenza nel caso di applicazione del principio “Open Data by default”.

<https://www.comune.palermo.it/xmls/VIS_DATASET_TURISMO03.xml>

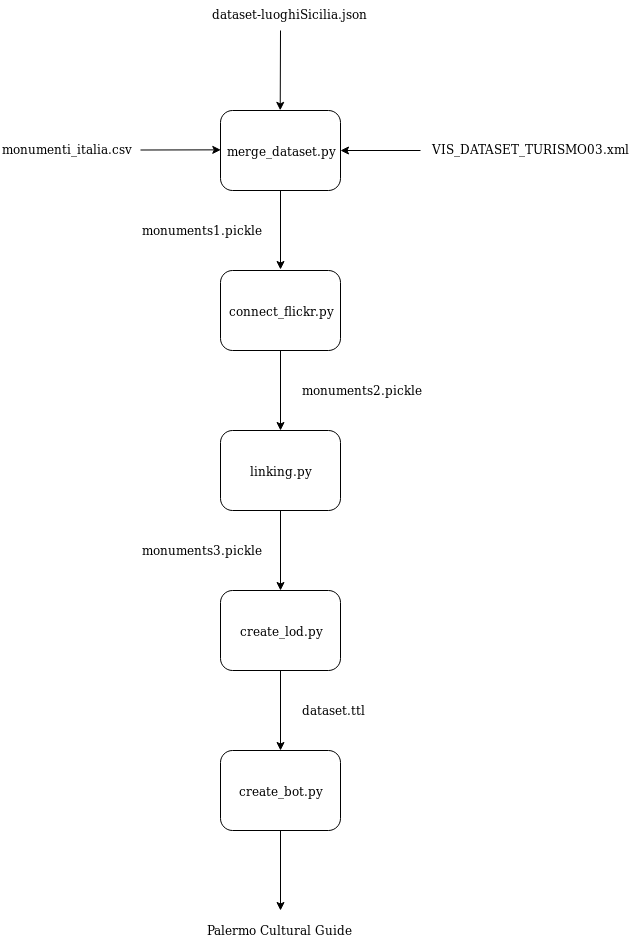
Licenza: ODBL:

Licenza che permette di condividere, adattare e creare il database, a condizione di attribuire, condividere allo stesso modo e mantenere aperto il database.

[www.datiopen.it/export/csv/Mappa-dei-monumenti-in-Italia.csv](http://www.datiopen.it/export/csv/Mappa-dei-monumenti-in-Italia.csv)

3. Definizione ontologia

4. Pipeline elaborazione



N.B.

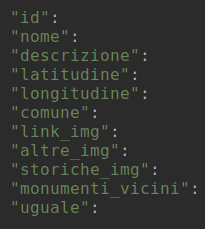
Durante tutta la pipeline di elaborazione, per poter utilizzare la struttura dati monuments in più codici sorgenti, è stata utilizzata la libreria pickle.

4.1 Unione dei dataset

In questa prima fase di elaborazione, è stata realizzata l’unione di tutti i dataset proveniente da diverse fonti (citate precedentemente). In particolare, è stata definita per ogni tipo di file (csv, xml, JSON) un’apposita funzione.



Tutti i siti culturali vengono raccolti all’interno di una lista di dizionari, aventi come chiavi:

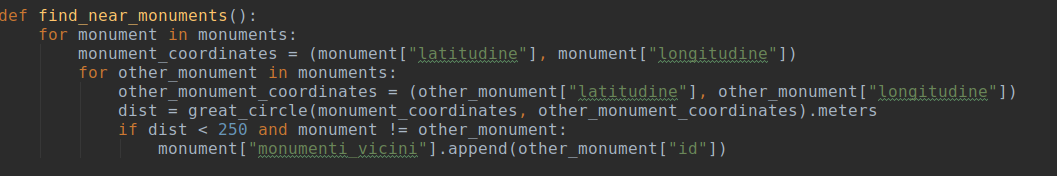


Ovviamente, è stato necessario apporre determinati controlli per ogni tipo di file analizzato. Nello specifico, avendo aggiunto per primi i dati ricavati dal file xml, poichè il file più completo, durante l’estrazione delle informazioni dagli altri due file, è stato necessario evitare l’aggiunta di eventuali duplicati. Per far ciò è stata utilizzata la libreria difflib (in particolare il modulo SequenceMatcher) per effettuare un controllo sulla similarità dei nomi dei siti culturali.

FARE ESEMPI

Quindi è stato sperimentalmente osservato che l’indice di similarità appropriato fosse di 0.8.

Successivamente per poter identificare i siti culturali vicini ad un altro, è stata sviluppata la seguente funzione che fa uso della libreria geopy.distance:



4.2 Unione dataset con flickr

Per poter aggiungere al nostro dataset le immagini messe a disposizione dalla biblioteca comunale di Palermo, all’interno del loro account flickr, abbiamo utilizzato le API fornite da flickr.

Attraverso una richiesta HTTP GET ed utilizzando il metodo flickr.people.getPhotos, flickr restituisce un file JSON. Per esempio:

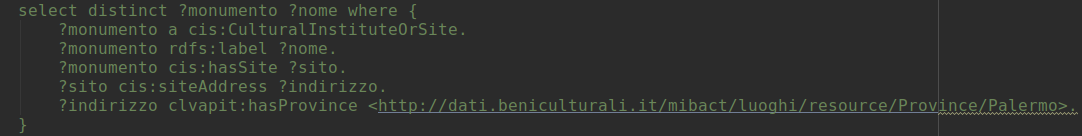


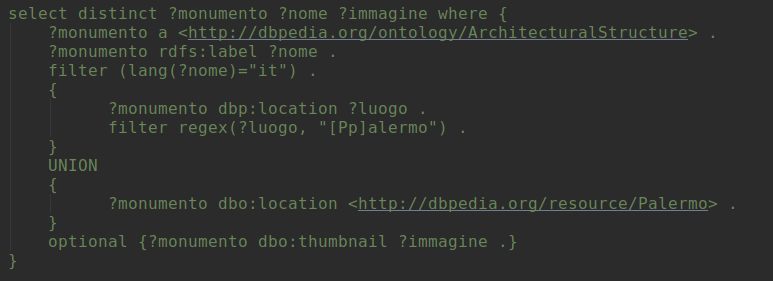
Questi dati restituiti, sono stati manipolati in maniera opportuna per attribuire ad ogni sito culturale, la relativa immagine. In particolare, come precedentemente, per effettuare l’associazione tra i siti culturali presenti all’interno della nostra struttura e le immagini, è stato riutilizzato il modulo SequenceMatcher. Anche questa volta, l’indice di similarità, è stato impostato a 0.80 dopo diverse prove sperimentali.

FOTO PROVE

4.3 Interlinking

Per rendere il nostro dataset linked, la nostra base di conoscenza è stata collegata alla base di conoscenza di dbpedia e dati.beniculturali.it . Per realizzare ciò sono state effettuate query SPARQL ai due endpoint.





Anche questa volta, per realizzare l’associazione tra i nostri siti culturali e quelli delle due basi di conoscenza, è stata effettuata una verifica sulla similarità dei nomi. ………..

FARE FOTO SPERIMENTALI