



Popular Hashtags around the word

Prepared by: Francesco Cardascia

Prepared for: Laboratorio di Progettazione Web A.A. 2015-2016

15 Febbraio 2016

DESCRIZIONE

Obiettivo

Il progetto nasce come sviluppo per l'esame di laboratorio di progettazione web presso l'università di Pisa. Il fine dell'applicazione è quello di mostrare la popolarità e la diffusione degli hashtag più utilizzati sulla piattaforma social 'Instagram' durante un determinato periodo di tempo relativo alla raccolta dei dati. Per studiare al meglio le dinamiche che determinano la scelta dell'accostamento degli hashtag da parte degli utenti, l'applicazione abbina alla lista dei dieci hashtag più popolari sulla piattaforma quella della frequenza delle co-occorrenze relative ad ogni specifico hashtag 'master'.

Fase preliminare - Raccolta dati

Durante la fase preparatoria di raccolta dei dati, è stato necessario effettuare lo scraping di un sito che si occupa di stilare liste di frequenza sull'uso di parole precedute dal cancelletto sulle varie piattaforme di interazione social. Nella fattispecie sono state ricavate le dieci parole trending su Instagram e salvate (con il relativo numero di frequenza) in una tabella nel database.

Successivamente è stato settato un *cron job* che ogni mezz'ora, per la durata complessiva di circa un mese e mezzo, ha raccolto mediante le API di Instagram (endpoint tags) le ultime 20 foto di ognuno dei 10 hashtag popolari. Di queste, sono state analizzate e salvate soltanto quelle geolocalizzate, da cui sono state estratte le seguenti informazioni: url, descrizione, hashtags, latitudine, longitudine e media_id. A partire da latitudine e longitudine relative ad ogni singola immagine sono state interrogate le API di Foursquare (endpoint venues) al fine di determinare la nazione di appartenenza della foto.

#	Name	Type	Collation	Attributes	Null	Default	Extra	Action
<input type="checkbox"/>	1 ID	int(11)			No	None	AUTO_INCREMENT	Change Drop More
<input type="checkbox"/>	2 media_id	varchar(100)	utf8_general_ci		No	None		Change Drop More
<input type="checkbox"/>	3 hash	varchar(20)	utf8_general_ci		No	None		Change Drop More
<input type="checkbox"/>	4 img	varchar(200)	utf8_general_ci		No	None		Change Drop More
<input type="checkbox"/>	5 url	varchar(200)	utf8_general_ci		No	None		Change Drop More
<input type="checkbox"/>	6 caption	varchar(300)	utf8_general_ci		No	None		Change Drop More
<input type="checkbox"/>	7 tags	varchar(900)	utf8_general_ci		No	None		Change Drop More
<input type="checkbox"/>	8 lat	varchar(30)	utf8_general_ci		No	None		Change Drop More
<input type="checkbox"/>	9 lng	varchar(30)	utf8_general_ci		No	None		Change Drop More
<input type="checkbox"/>	10 country	varchar(50)	utf8_general_ci		No	None		Change Drop More
<input type="checkbox"/>	11 country_code	varchar(10)	utf8_general_ci		No	None		Change Drop More

Fase intermedia - Elaborazione

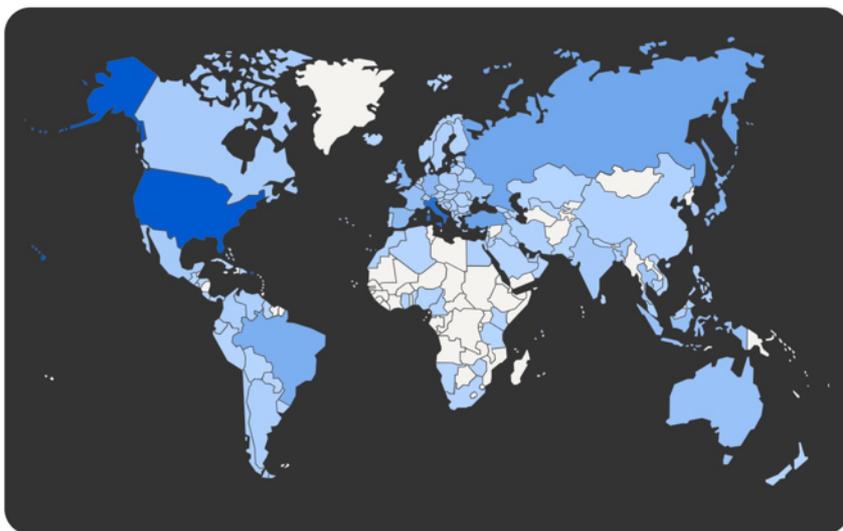
Nel database ogni record relativo a ciascuna foto presenta (come sopra indicato) un campo 'tags'. È stato salvato il contenuto di quel campo (raggruppando per hashtag) in file di testo singoli. Per ricavarne le frequenze è stato utilizzato Python, in quanto il tempo di timeout di esecuzione del php non era sufficiente ed in ogni caso, anche estendendolo, si i tempi richiesti non erano accettabili. I dati sono stati poi salvati in formato JSON.

Visualizzazione

Mediante l'utilizzo di apposite chiamate AJAX, al caricamento della pagina, i dati vengono serviti direttamente dal database che viene interrogato tramite delle semplici query mysql. L'uso della libreria jQuery, oltre ad aver reso possibili le suddette chiamate, rende possibili tutte le interazioni con la pagina. Si è scelto di strutturare il sito in un'unica pagina HTML per rendere l'interazione dell'utente il più lineare e semplice possibile.

L'interfaccia presenta una lista dei dieci hashtag "chiave" che, una volta selezionati, innescano il caricamento dei dati relativi alla diffusione su scala mondiale dell'hashtag selezionato.

#fashion
#friends
#smile
#like4like
#instamood
#family
#nofilter
#amazing
#style
#follow4follow



Nella sezione sottostante (*Distribution*) vengono mostrati gli hashtag co-occorrenti a quello selezionato. Viene generato quindi un tag cloud che presenta i risultati più importanti. Si è scelto di sfruttare questa specifica modalità di visualizzazione appunto per la caratteristica intrinseca della “nuvola di etichette”, nella quale ogni etichetta ha grandezza pari al suo peso, che corrisponde al numero di co-occorrenze.

Distribution



Cliccando il tasto “Show more”, invece, vengono mostrati via via tutti gli hashtag rimanenti in ordine decrescente.

Al click su ogni parola della cloud o della lista presente nella sezione “Show more” vengono caricate e visualizzate le immagini corrispondenti, con i relativi hashtag, caption e paese di provenienza.

STATO DELL'ARTE

Top-hashtags.com

Il sito presenta una lista (in continuo aggiornamento) degli hashtag più popolari sulle piattaforme social più diffuse quali: Instagram, Twitter, Tumblr. La lista ordinata per popolarità delle parole "di tendenza" è seguita da una seconda che riguarda il numero di frequenza con la quale esse occorrono. È un buon metodo per farsi un'idea indicativa dei *trends* di un determinato periodo, ma non fornisce un metodo che possa essere utile a sfruttare questi dati (se non quello di condividere l'hashtag stesso e contribuire alla crescita del trend). Il sito inoltre non dà nessuna indicazione circa i contenuti veri e propri e non dà la possibilità di interagirvi.

Partendo dal concetto base di top-hashtags.com si è cercato di realizzare un'applicazione che permettesse, non solo di tenere traccia dei trend in un determinato periodo di tempo, ma anche di darvi una collocazione spaziale. È infatti possibile farsi un'idea della distribuzione su scala globale di queste parole chiave e successivamente apprezzare le differenze significative che si vengono a creare tra un hashtag e l'altro.

Tramite la visualizzazione sulla mappa la differenziazione per scale tonali rende molto intuitivo e facilmente evidenziabile quali siano gli hashtag più utilizzati in determinati paesi piuttosto che in altri. Inoltre, anche il tag cloud ha il suo forte impatto visivo molto più indicativo di una semplice lista di parole che, per quanto emblematica, risulta comunque essere fine a se stessa.

In aggiunta, scorrendo le foto relative all'accoppiamento hashtag master con uno dei suoi relativi, facendo caso al paese di provenienza, in alcuni casi, è possibile fare una sorta di indagine sociologica per poter avanzare delle ipotesi di quali siano gli atteggiamenti delle persone in tutto il mondo riguardo l'argomento social network in primis, riguardo Instagram in questo caso, il suo utilizzo, la classificazione in base agli hashtag ed il loro (ab-)uso, coerentemente con i servizi di geolocalizzazione di ciascuna foto.

Inoltre è stato possibile evidenziare che l'hashtag, ormai, ha la funzione inversa rispetto al motivo e la filosofia per cui era nato. Anziché raggruppare per argomento, in questo caso, le foto è divenuto un tramite per raggiungere più gente possibile. Confermano questa ipotesi hashtag come "like4like", "follow4follow" o "instamood" e via dicendo che sono inseriti di routine dagli utenti soltanto per accaparrarsi più like e quindi più followers.
