

TITOLO: Un approccio open source per la valutazione intrinseca di accuratezza tematica, accuratezza temporale, aggiornamento e lignaggio di OpenStreetMap

AUTORI: Francesco Frassinelli (1), Marco Minghini (1), Maria Antonia Brovelli (1)
(1) Politecnico di Milano, Department of Civil and Environmental Engineering, Milan (Italy)
marco.minghini@polimi.it, maria.brovelli@polimi.it, francesco.frassinelli@mail.polimi.it

KEYWORDS: Aggiornamento, Accuratezza tematica, Accuratezza temporale, Lignaggio, OpenStreetMap, Qualità dei dati

ABSTRACT:

Iniziato ben 13 anni fa, OpenStreetMap (OSM, <https://www.openstreetmap.org>) è oggi senza dubbio il progetto di Informazione Geografica Volontaria di maggior successo. Nonostante i dati di OSM siano attualmente utilizzati da moltissimi soggetti e per una miriade di applicazioni diverse, il problema principale che ne limita un utilizzo ancora più esteso è quello - tipico dei progetti di crowdsourcing - della qualità. La letteratura ha proposto finora innumerevoli approcci per valutare, in modo più o meno automatico e generalizzabile, i diversi parametri di qualità per i dati geospaziali: accuratezza posizionale, completezza, accuratezza tematica, accuratezza temporale, aggiornamento e lignaggio. Tali approcci si classificano in due categorie: estrinseci, se la valutazione della qualità avviene per confronto con un dataset esterno di riferimento (ad esempio una cartografia ufficiale) ed intrinseci, se la valutazione della qualità avviene analizzando unicamente i dati stessi di OSM.

Il presente lavoro si inserisce in questo contesto per sopperire alla quasi totale assenza di approcci intrinseci per la valutazione (sia separata che congiunta) dei parametri di qualità di accuratezza tematica, accuratezza temporale, aggiornamento e lignaggio. In primo luogo, sono state create alcune mappe web che, a partire dal database di OSM e per qualsiasi area selezionata dall'utente, offrono le seguenti visualizzazioni inerenti ai parametri di qualità sopra citati: mappa che mostra la data di ultima modifica di ogni nodo e way; mappa che mostra la data di creazione di ogni nodo e way; mappa che mostra il numero di versione (ovvero di changeset) di ogni nodo e way; mappa che mostra la frequenza di aggiornamento di ogni nodo e way; mappa che mostra il numero di utenti diversi che hanno editato ogni nodo e way. Tali mappe, che hanno già ricevuto feedback positivi dalla comunità italiana di OSM, sono molto utili per fornire un'indicazione immediata su quali dati potrebbe essere necessario o prioritario aggiornare. A titolo esemplificativo, la prima delle mappe elencate è accessibile alla pagina <https://is-osm-uptodate.frafra.eu>. Per ciascuna di tali mappe, infine, è in corso di sperimentazione la possibilità di una visualizzazione aggregata tramite griglia, le cui celle siano visualizzate in scala cromatica come risultato di un algoritmo basato sulla quantità, densità, frequenza ed ultimo aggiornamento delle modifiche dei nodi e way presenti nella cella.

In aggiunta alla semplice visualizzazione, le numerose analisi svolte a partire dalla storia del database di OSM (in prima battuta limitate al caso italiano) consentono di trarre qualche conclusione generale. In primo luogo l'accuratezza temporale, l'aggiornamento e il lignaggio dei dati OSM sono in generale maggiori nelle aree urbane (più "interessanti" da mappare e dove sono attivi molti contributori) rispetto a quelle rurali. Ciò è in accordo con i risultati tipicamente ottenuti in letteratura per il parametro della completezza. In ogni caso, sebbene il trend globale sia piuttosto delineato, a livello locale si nota una disomogeneità dei risultati tipica di OSM e dovuta all'effettiva presenza e attività dei contributori che, talvolta, possono essere molto diverse anche tra città e comuni confinanti.

Una seconda analisi ha invece studiato l'accuratezza tematica, ovvero degli attributi (o tag) associati ai nodi e alle way di OSM, ed il suo trend temporale per diverse categorie di oggetti (attività commerciali,

elementi di arredo urbano, elementi naturali, ecc.). Per tali categorie, sono state individuate alcune ricorrenze tipiche nell'ordine di aggiunta dei tag e definiti trend che, in base al numero di tag esistenti e allo storico della loro mappatura, consentono un'indicazione automatica della loro accuratezza tematica. L'analisi è svolta a partire dall'intero file Planet/Full di OSM, che contiene sia tutti gli oggetti geografici presenti nel database, sia la loro evoluzione temporale. Le elaborazioni sono interamente eseguite con software open source. In particolare il file Planet è convertito in un database SQLite grazie a una procedura scritta in Python che utilizza i bindings della libreria Osmium, mentre le mappe web menzionate precedentemente sono realizzate con Leaflet e Bootstrap per la parte di frontend e con Python 3, SQLite, SpatiaLite, Spatialite-tools, Hug, e uWSGI per la parte di backend. Le analisi richiedono di affrontare problemi computazionali legati a gestione di grandi basi di dati, in particolare alla conversione del Planet/Full in un database SQL, aggravata dalla scarsità di strumenti e di librerie in grado di gestire la history dei dati di OSM; di conseguenza differenti approcci al problema sono al momento in fase di test.