

# Geoblogging - User-contributed Geospatial Data Collection and Fusion

Dieter Pfoser  
Charikleia Lontou  
Euthymios Drymonas  
Institute for the Management  
of Information Systems  
G. Mpakou 17  
11524 Athens, Greece  
pfoser|clontou|edrimon@  
imis.athena-innovation.gr

Savvas Georgiou  
School of Electrical and  
Computer Engineering  
National Technical University  
of Athens  
15780 Athens, Greece  
el04254@mail.ntua.gr

## ABSTRACT

One cannot deny that space and time are important to us. We perceive our world with respect to where and when we do things. We advocate geoblogging as a tool to capture such experiences by means of collecting and organizing notes, images, and in relation to space and time as well as to link them to other geospatial datasets. This demo showcases a Web application that allows for (i) a simple upload of content, geocoding, and map-based authoring of geoblogs as well (ii) querying and linking other geospatial datasets in relation to the geoblog entry.

## Categories and Subject Descriptors

H.5.3 [Information Systems Applications]: Group and Organization Interfaces, Web-based interaction

## Keywords

geoblogging, data fusion, user-contributed content

## 1. INTRODUCTION

When coming home from a memorable journey, wouldn't it be great to create a digital replica of the trip, i.e., quickly organize collected images and videos, discover and link additional content from the Web (e.g., Flickr images, restaurant reviews) and have a simple means to group all this content and add some thoughts? Should you be using Web 2.0 services such as Flickr, Picasa, etc. you might notice that such applications are data oriented. They provide for an efficient means of storing a specific type of content, including some limited content management functionality such as geocoding and grouping.

With our application, termed "Geocrowd", we propose *geoblogging* as a means for *spatiotemporal storytelling*, more specifically the story of a journey, be it an afternoon walk in your neighborhood, a chase for a great coffeeshop, your mountainbike trip, or hiking adventure. The scope is to provide a simple to use application that

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM GIS '10 November 2-5, 2010, San Jose, CA, USA  
Copyright 2010 ACM ISBN 978-1-4503-0428-3/10/11 ...\$10.00.

allows one to tell the story of the trip based on the content (i) collected during the trip as well as (ii) data discovered on the Web. Hence, Geocrowd is a means to *aggregate geospatial data* with respect to specific experiences. In our application, the essential aspects are a map, a storyboard and a timeline. Overall, we see geoblogging as a means to harness the ability humans have to massively *collect, fuse and share knowledge* ultimately digitizing the world using user-contributed content.

While no specific work exists towards the present idea of geoblogging, related approaches and applications include typical photo sites such as Flickr or Panoramio, sites for collecting trail information, e.g., trailguru [4]. Using a GPS-equipped camera taking continuous snapshots, geoblogging can also be found as an aspect of the MyLifeBits project aimed at recording all information related to a person in digital form [1]. The application closest to our approach is Everytrail [2], which collects trail information in the form of photos and allows one to add a story to each trail. However, no actual geospatial knowledge collection is supported, i.e., trails and images are collected, but there is no specific editing functionality nor does the application provide means for linking external data. The remainder of the demo paper surveys the functionality and architecture of *Geocrowd*, our geoblogging application.

## 2. GEOBLOGGING FUNCTIONALITY

The geoblogging application is implemented as a simple to use Web application. Its *core functionality* includes content upload, geocoding and authoring, i.e., the actual geoblogging activity. *Extended functionality* provides a simple means to link external geospatial data sources. External sources are queried in relation to a geoblog entry using available APIs.

The Geocrowd service has been implemented as a Ruby-on-Rails/PostgreSQL Web application using Javascript to implement browser functionality. Besides the functionality described in the following, various interfaces and mobile clients have been implemented. Geocrowd data can be *exported* as a kml file that can be viewed in, e.g., Google Earth. Two mobile client applications have been developed for the Android platform; a live blogging client allows for the collection of data and direct submission to the service with routes being created on-the-fly and an augmented-reality (AR) application provides a mobile interface for the consumption of routes created in Geocrowd. A Geocrowd API has been implemented to connect to the service.

### 2.1 Core Functionality

The core functionality includes an upload interface, drag-and-drop geocoding and storyboard-based blogging functionality. The *upload interface* utilizes html5 multi-selection and, if not supported, uses a simpler interface supported by all browsers.



Figure 1: Editing text and geocoding

An important aspect for geoblogging is the geospatial position of the content. Supported *geocoding* means are (i) geotags as part of the content metadata (from GPS-equipped cameras), (ii) time-stamp matching using tracking data (GPS logger, GPX files) and (iii) drag-and-drop geocoding using a map, i.e., dragging an item to a location on the map (cf. Figure 1). The latter allows us to change any geocoding, e.g., to locate depicted objects rather than the location the picture was taken from.

Having uploaded all the content, the actual *geoblogging* is supported by a dual-view window consisting of a *storyboard* and a *map*. The storyboard can be seen similar to a text processing tool with the important difference that the vertical dimension implements a timeline and facilitates the chronological ordering of content, i.e., items on top have been visited before items at the bottom (cf. Figure 1). Text such as explanations, image captions, etc. can be added here and geocoded using the map view. Grouped items are indicated by an orange frame (cf. Figure 1). Grouping is a powerful operator that essentially supports the creation of places, i.e., linking various geospatial information pieces together. The storyboard uses a grid to simplify content (re-)ordering and placement, i.e., inserting a text automatically rearranges the content. The *map* serves two purposes, (a) it supports geocoding by dragging content from the content pane onto the map and (b) it shows the content arranged according to a route representing the trip. Using the chronological ordering in the storyboard and the actual positions on the map, a *route* connects all content items. Each segment of the route can either be a straight line (default) or a route in the road network utilizing the routing service of the Google Maps API. Routing is dynamic in that any change to the chronological order and/or position will immediately be reflected in an updated route. Following the above steps, the user is left with a geoblog entry such as the example shown in Figure 1.

## 2.2 Geodata APIs and Data Fusion

To provide the user with more content and richer blog entries, Geocrowd allows for the integration of third-party data sources (*linked content*). Linked content comprises point-of-interest (POI) data, i.e., content (pictures, descriptions, reviews) related to a location. Typical examples are Lonely Planet (general POI data), Museella (museum info), Qype (POI reviews/ratings), Wikipedia (POI info) and more recent sources such as Foursquare, Loopt, etc. While the ex-

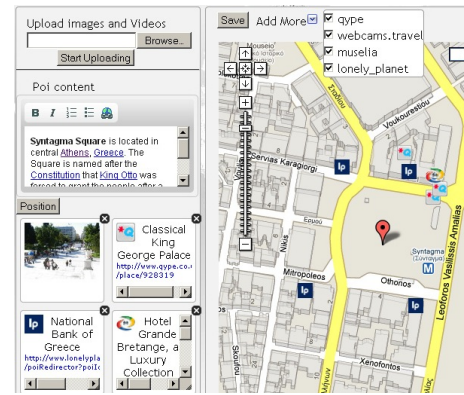


Figure 2: Creating POIs

amples so far represent *curated datasets*, i.e., each entry refers to a specific location/place that typically appears only once in each dataset, we also encounter *un-curated* data. The most prominent examples are Flickr and Panoramio, with each containing thousands of pictures for popular locations (cf. Google Maps view with Photos enabled). To link these sources, we created a filter that automatically derives POI datasets from such user-contributed content using lexical and spatial clustering. Lexical clustering employs an IE technique comprising three steps, (i) the preprocessing step, recognizing and extracting multi-word terms from retrieved image labels using the *C/NC*-value method [3], (ii) the indexing step, storing the texts as vectors of terms, each with an associated *tf-idf*-derived weight, and (iii) the clustering step that groups similar pictures based on their respective term vectors. Spatial clustering then computes the location of each retrieved POI based on its associated point cloud.

Linked content is retrieved for a region around a given route/POI. Excerpts of the content are stored locally along with a time stamp to assess data freshness and so to avoid redundant API requests in case of overlapping (in time and space) queries.

Linked content is fully-integrated in the editor as it can be added to a geoblog entry by dragging it from the content pane to the storyboard (cf. Qype Hotel item in Figure 1). Here grouping becomes a powerful feature since we essentially link data from various sources, representing a first step towards an integrated geospatial dataset. To underline the data fusion aspect, Figure 2 shows the POI creation functionality (Syntagma Square, Athens) relating data from various linked content sources in addition to one's own.

## 3. DEMO DESCRIPTION

The demo will showcase the geoblogging application and demonstrate the various features as described in the previous sections. Users will be able to create their own geoblog entries and try out the application using their own laptops/computers right away. The prototype of the geoblogging service is available at <http://www.geocrowd.org>. Since this is a project in progress, some of the described functionality may change over time.

## 4. REFERENCES

- [1] G. Bell and J. Gemmill. A Digital Life. *Scientific American*, 296:58–65, 2007.
- [2] Everytrail - GPS Travel Community. [Online]. Available: <http://www.everytrail.com/>.
- [3] K. Frantzi, S. Ananiadou, and H. Mima. Automatic recognition of multi-word terms: The *c*-value/*nc*-value method. *Intern. Journal of Digital Libraries*, pages 117–132, 2000.
- [4] Trailguru. [Online]. Available: <http://www.trailguru.com/>.