Collecting life logs for experience-based corpora

F. Francesconi, A. Ghosh, G. Riccardi, M. Ronchetti and A. Vagin

DISI - University of Trento, 38050 Povo (Trento), Italy

{francesconi, arindam.ghosh, riccardi, marco.ronchetti, alex}@disi.unitn.it

Abstract

In this paper we propose an approach to lightweight acquisition, sharing and annotation of experience-based corpora via mobile devices. Corpora acquisition is the crucial and often costly process in speech and language science and engineering. To address this problem, we have built a system for creating a location based corpora annotated with multi-media tags (e.g. text, speech, image) generated by end-users. We describe a relevant case study for the collection of mobile user life logs. We plan to make publicly available such tools and platforms to the research community for collaborative development and distributed experiential corpora collection.

Index Terms: corpora creation, mobile data collection.

1. Introduction

Corpora acquisition is the crucial and often costly process in speech and language science and engineering. Most of the available corpora are primarily spoken or written text databases from diverse sources. Constructing a large language corpus has always been a complex and expensive task requiring careful planning, elaborate setup and dedicated infrastructure. Hence researchers have often had to depend on centralized and costly process such as in Switchboard [1] and ATIS [2]. More recently, live data collections engaging real users may be either recorded under controlled conditions or proprietary (e.g. contact centers).

To overcome these limitations, there have been attempts to perform speech data collection in the natural habitat of a user. Most of these experiments require the user to carry some kind of portable recorder, which records the user's speech and actions. In one such experiment, the Electronically Activated Recorder or EAR [3], attempts to periodically record small pieces of ambient sounds using a wearable device carried by a user. In another recent experiment conducted by Google, [4] a transcribed speech corpora collection tool on the Android smart-phone platform was developed. This tool provides textual prompts for the speaker to read. It collects data related to speaker's demographics and GPS location. However such experiments have their disadvantages. The data collected tends to be non-spontaneous, of a single type of media (either speech or video), and does not have the other contextual information of the user augmented with it. Speech and language research would benefit from contextual information regarding the speakers' experience including the language traces and behavioral signals.

We hereby propose a system to collect, share and annotate multimedia data, which can be further augmented with user's experiences and other contextual information. Such an experiential corpus can be described as a consistent set of events over time in a physical world where we want to observe signals (verbal/non verbal) from the user and his activities. An event can be either active, in the form of a user recording a voice message, typing some text, or taking a photograph; or more covert or passive where a user's physical state such as location, heart-rate or body temperature changes are recorded. A few years back such a complex data collection task would have been a daunting task. However the recent improvements in the smart-phone technology has opened up a new avenue for augmenting user experiences with collected speech and text corpora. Using a smart phone a user can effortlessly record voices, take photographs or log messages in the form of text. The smart phone can also record the user's location, provide data about the user's movement using the accelerometer and communicate with other sensors on the user's body to understand his complete experience.

Motivated by these considerations, we have built the iScout system: a platform for creating a location based multi-media corpora. iScout currently includes an iPhone application for collecting speech, text, image and location data while users travel or visit different places; and an online platform Giscover for creating a collaborative platform around this collected data. Users can record their daily tracks, annotating them with speech, text and images, and later view it offline or upload it to the Giscover community to share it. Other people following similar trails during trekking or touristic visits can view the data on this community. The user-annotated files are available in KML format [5] for later manual or automated processing (e.g. metadata annotation /augmentation).

2. System Architecture

Our current implementation of the iScout system is based on client/server architecture, composed by an iPhone app on the client side, and a web application called Giscover [6] on the server side.

2.1. The Client system

The client system is currently implemented as an Apple iPhone app running on the iOS platform to record and store speech and other multimedia data collected by the user. The phone also provides additional metadata such as geographical position of the speaker, which can be augmented with the data collected.

The iScout platform uses the paradigm of tracks. The user can enable the tracking-mode of the iScout application, which continuously tracks the user's geographical position and also displays it on the application's internal map. The user can add a waypoint at any geographical location and can enrich it by adding audio messages, images, or text notes which are immediately annotated with the user's information.

iScout makes use of the Core Location framework of the iOS to retrieve location information from the phone's GPS and draw real-time trails on the map whenever the user's location gets updated. A limitation of the Google map on iOS is that it does not allow updating of already existing routes on the map. As a solution an aggregation algorithm [7], which keeps merging chunks according to a specific aggregation factor was used. iScout can run as a background task, so that even when the user minimizes the application or locks his phone the GPS location continues to be tracked by the iScout application thus tracing the path.

Figure 1**:** *Screenshots of iScout*

2.2. The Server side

The Server is based on the Giscover platform, a system that provides the users a unique set of location-based services integrated with social network interactions. The Giscover platform allows the end-users to upload the data collected using iScout to their website to create a "tour-composition". A tour composition is a collection of media contents assembled from a number of items such as photographs, audio and video recordings, points of interests (POIs), and location based GPS trails.

Figure 2**:** *iScout and Giscover*

Users can also create channels consisting of a collection of tours. A Giscover Channel [8] is a unique feature incorporating a set of pages that users can customize the look and feel of and later embed it into their web sites, export or link data to other services using KML (Keyhole Markup Language). Additionally, channels as well as single tours can be shared across social networks, such as Facebook, Twitter and Google Buzz. Using the Giscover platform the user can enrich and ease access to their content in several ways, namely:

- Automatic geo-tagging of photos and video and propagation back to the original service;
- Tour animations and KML interactive editor with features like "drag & drop";
- Audio recordings and their conversion to formats supported by the most of modern browsers;
- Reverse geo-coding and reverse geo-discovery;
- QR code integration;
- Post editing and annotating of tours by users;
- Comment on the users' tours by the community members.

3. Discussion and conclusion

The iScout system has already been used to collect a large body of data in the domain of trekking or city touring. We have also experimented with iScout in the shopping scenario. By opening the tools and platform to the research community we expect to enrich it with annotation tools, sensors and interfaces. The goal is to make iScout and its collaborative developments a research vehicle to collect experiential corpora.

4. Acknowledgements

The research has been partly funded by the Livememories project funded by the Province of Trento (Italy). We would like to thank Belén Agüeras Angulo for her support on the development of the app UI.

5. References

[1] John J. Godfrey and Edward Holliman. 1997. Switchboard-1 Release 2. Linguistic Data Consortium, Philadelphia.

[2] Hemphill, C. T.; Godfrey, J. J. & Doddington, G. R. "The ATIS Spoken Language Systems Pilot Corpus", In Proceedings of the DARPA Speech and Natural Language Workshop, Morgan Kaufmann, 1990, 96-101.

[3] Fisher, K.; Henary, B.; Mcfarl, J.; Orr, K.; Shook, N.; Mehl, M.; Pennebaker, J. W.; Crow, D. M.; Dabbs, J. & Price, J. H. The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations, 2001.

[4] Hughes, Thad / Nakajima, Kaisuke / Ha, Linne / Vasu, Atul / Moreno, Pedro J. / LeBeau, Mike (2010): "Building transcribed speech corpora quickly and cheaply for many languages", In Interspeech-2010, 1914-1917.

[5] KML http://www.opengeospatial.org/standards/kml

[6] Giscover http://www.my.giscover.com/

[7] Fabiano Francesconi, Walter Da Col. MapKitDynRoutes https://github.com/elbryan/MapKitDynRoutes

[8] Giscover Channels http://my.giscover.com/channels/6-live-memories