The Field Linguists' App

1 Project Abstract

The FieldLinguists' App is an OpenSource database that allows language researchers to securely enter, store, organize, annotate, and share linguistic data. The application is accessible everywhere; it runs on three different systems (Mac, Linux, and Windows) and is suitable for both online and offline use. Furthermore, the application will be created with collaborative goals in mind; data will be syncable and sharable with other researchers. Researchers can form teams that contributes to a single corpus, where team members can modify and discuss the data from within the application. The system will also have a simple and friendly user interface, allowing participants to drag and drop files (audio, video, text), or record it directly into the database when using the Android app. In addition, the application will have import and export capabilities for multiple file types. Most importantly, the application is designed intuitively and theory free, so it is not necessary to be a field linguist or programmer to figure out how it works.

2 Statement of Need

The statement of need should describe the problem that the project will attempt to address. Also, describe the population that will be served.

3 App Description

Describe the project or program and provide information on how it will be implemented. Include information on what will be accomplished and the desired outcome.

4 Goals & Objectives

Describe the project objectives in measurable terms that address the academic and technological needs of language teachers, linguists, etc. (it's for linguists but other people

can benefit from it as well i.e. help create dictionaries for endangered languages, which will benefit the communities of these languages; L2 acq. teachers).

5 Staff & Organizational Information

iLanguage Lab is a Montreal based company that develops tools in the form of experimentation and data collection apps for Android and Chrome in collaboration with researchers at UdeM, UQAM, McGill and Concordia. Previous research applications includes the Bilingual Aphasia Test, AuBlog, OPrime and SpyOrNot. Furthermore, iLanguage Lab has a background in assisting researchers obtain results and publications. The AuBlog application was employed to investigate evolving information structure and audienceless vs. audience oriented prosodies and culminated in a poster presented at Experimental and Theoretical Advances in Prosody Conference. The Bilingual Aphasia Test led to a presentation at the Academy of Aphasia 49th Annual Meeting on Aphasia Assessment on Android: recording voice, eye-gaze and touch for the BAT and a publication in the Academy of Aphasia.

5.1 Gina Cook M.A.

Gina Cook received her Masters in Field Linguistics & DESS in Computer Science and has worked as a computational linguist for companies such as Nuance and Idelia. She founded iLanguage Lab with a vision to develop computational tools to help researchers as opposed to consumers. She is an active contributor to OpenSource projects on GitHub focusing on integrating existing OpenSource libraries for Speech Recognition, Natural Language Processing, Eye Gaze analysis and Acoustic analysis into Android tablet applications.

5.1.1 Publications

- "Aphasia Assessment on Android: recording voice, eye-gaze and touch for the BAT." (with A. Marquis & A. Achim). Poster at Academy of Aphasia 49th Annual Meeting, Montéal, Québec. October 2011.
- "Eliciting evolving information structure and audienceless vs. audience oriented prosodies: experimentation on Android tablets." (with S. Kattoju). Poster at ETAP2 Ð Experimental and Theoretical Advances in Prosody, Montréal, Québec. September 2011.
- "PDFtoAudioBook Android app" (Java, XML).Canadian University Software Engineering Conference (CUSEC) DemoCamp, Montréal, Québec. January 2011.

- "Word features and word concatenation."Sixth Interdisciplinary Graduate Student Research Symposium, McGill University, Montréal, Québec. March 2009.
- "The Structure of Long Distance Agreement in Hindi/Urdu." Invited Lecture in Advanced Syntax, Concordia University, Montréal, Québec. November 2007.
- "The Phonological/Phonetic status of Productive Palatalization in Romanian." (with L. Spinu). Presented at the Seoul International Conference on Linguistics, Seoul National University, Seoul, South Korea. July 2006.

5.2 M.E. Cathcart Ph.D.

M.E. Cathcart completed her PhD at the University of Delaware with a dissertation grant funded by the National Science Foundation (NSF) for her field work in Cusco, Peru on Quechua. In addition, she also has a background of coursework in computational linguistics, at the University of Delaware and at the Linguistic Society of AmericaÕs Summer Institute.

5.2.1 Publications

- "Affected Arguments Cross-linguistically." S. Bosse, B. Bruening, M.E. Cathcart, A. E. Peng, M. Yamada. In: Tadic, M. Dimitrova-Vulchanova, M., Koeva, S. (eds.): FASSBL 6 The Sixth International Conference on Formal Approaches to South Slavic and Balkan Languages. 2008 (Proceedings) pp. 41-47.
- "A New Grammatical Category: Impulsatives." Penn Linguistics Colloquium, Philadelphia March 2010 ?Eliciting data for dissertation on Impulsatives: functional morpheme in context
- "The Syntax and Semantics of Desideratives in Albanian." Georgetown Linguistics Society, Washington, D.C. February 2010
- "Bi-Eventivity & Affecting Arguments." S. Bosse, B. Bruening, M.E. Cathcart, H.-j. Cheng, A. E. Peng, M. Yamada. Formal Approaches to South Slavic and Balkan Languages, Dubrovnik (Croatia), 25-28 September 2008.
- "Bi-Eventive Affect." S. Bosse, B. Bruening, M.E. Cathcart, H.-j. Cheng, A. E. Peng, M. Yamada.TEAL, Potsdam (Germany), 10-11 September 2008.

5.3 Theresa Deering M.A.

Theresa Deering has a Bachelor's in Computer Science from Malaspina and a Master's in Computer Science from McGill University. Her thesis focused on the Least-Used Direction pivot rule for the Simplex Method of solving linear programs.

- "The Least-Used Direction Pivot Rule on Acyclic Unique Sink Orientations." Master's Thesis. McGill University, Montréal, Québec. July 2010.
- "Worst-case Behaviour of History Based Pivot Rules on Acyclic Unique Sink Orientations of Hypercubes." Y. Aoshima, D. Avis, T. Deering, Y. Matsumoto, S. Moriyama. Submitted to AAAC. October 2011.

6 Budget

The Field Linguists' Database is composed eight modules and thus the cost is divided into eight major components. In addition, a separate price is given for software architecture and for 1 year of user support, and project growth; needed to make a longterm viable and useful tool that fieldlinguists can adopt for their labs or for their field methods courses. The collaboration module is used to permit collaboration with teams and users. The Corpus Module is used to sync, share, edit, tag, categorize and open data. The Lexicon Module is used to house, and read lexicon entries to be used for the glosser. The Dictionary Module is used to share the lexicon in the form of a WordNet/Wiktionary dictionary with the language community as required by some grants. The Glosser Module is used to automatically gloss datum, smarter than the standard lexicon. Finally, the Aligner Module is used to create TextGrids from the orthography and the audio files, used for prosody and phonetic analysis.

The prices given include a 40% Researcher discount and a 10% Open Source discount. An additional 40% discount is given for certain components that can be made with the help of students from Concordia University, Hisako Noguchi and Yulia Manyakina. The cost is calculated by determining the time in hours and multiplying by $48, the average rate of a software developer in Montréal.

6.1 Collaboration Module

Module	Price
Software Architecture	$1,555.20
Collaboration Module	$5,728.32
Corpus Module	$9,201.60
Web Spider Module	$2,177.28
Phonological Search Module	$2,177.28
Lexicon Module	$7,340.54
Dictionary Module	$18,781.63
Glosser Module	$17,770.75
Aligner Module	$ 9,787.39
User Support	$30,246.70
TVS and TPQ	$10,833.32
Total	$83,176.04

Table 1: Cost Summary

Collaboration Module				
Iteration	Hours	Technology	Rate	Cost Closed So
Software Architecture Design	20	Software Engineering	48	960
Collaboration API on central server	30	Software Engineering	48	1440
Users Model	15	Javascript	48	720
Informant Model	15	Javascript	48	720
Team Model	15	Javascript	48	720
Bot Model	15	Javascript	48	720
User Activity Model	8	Javascript	48	384
Team Feed Widget	25	HTML5	48	1200
User list item Widget	16	HTML5	48	768
Team Preferences Widget	8	HTML5	48	384
User Profile Widget	8	HTML5	48	384
User Tests	30	Javascript	48	1440
Informant Tests	30	Javascript	48	1440
Team Tests	30	Javascript	48	1440
Android Deployment	15	Java	48	720
Chrome Extension Deployment	20	Javascript	48	960
Heroku Deployment	5	Integration	48	240
# of weeks with 3 full time personnel	2.5416666667			14640

Table 2: Cost Summary