INTERSPEECH 2013: Show & Tell Contribution

Lexee: A Cloud-Based Platform for Building and Deploying Voice-Enabled Mobile Applications

Dmitry Sityaev, Jonathan Hotz, Vadim Snitkovsky

Angel Incorporated, United States of America

dsityaev@angel.com, jhotz@angel.com, vadim@angel.com

Abstract

This paper describes a new platform called Lexee developed as part of a cloud-based Angel offering which allows to rapidly develop voice-enabled mobile applications. The solution allows to rapidly build application flows without any significant investments into software or extensive knowledge of programming. The Lexee SDK described in this paper gives developers a starting point to create their own multi-modal mobile applications.

Index Terms: mobile applications, personal assistants, SaaS

1. Introduction

Advances in Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) technology have now reached a point where it appears viable to have a high quality voice-machine interface with minimal recognition errors. At the same time, access to commercially developed speech technology has greatly increased and this is particularly exemplified by the proliferation of applications on consumer devices. Technologies like Apple's Siri and Google Voice have certainly contributed to the trend: more and more people are exposed to the ubiquitous speech technology on the mobile device and have made voice interaction a part of their routine interaction with technology. Through the combination of all these factors mentioned above, a new trend has emerged where companies offer a product commonly known as a "specialized personal assistant" (SPA), and what was once only seen in science fiction is now becoming a part of our everyday lives.

One of the challenges faced when developing these new types of mobile applications is how to quickly and effectively build an SPA without having to construct the complete application from scratch. The technology needs to be there: that, coupled with understanding how users interact with a multi-modal voice-enabled mobile application, led to the development of the Lexee platform. Lexee handles all the aspects of the technical architecture, including the interaction with the ASR and TTS services, as well as development and deployment of the mobile application, allowing developers to focus on the interaction between the user and application. The Lexee platform includes a powerful tool for building the interaction using a visual point-and-click GUI to design the application flow, and does not require any advanced programming skills.

Developing and deploying an application is only the first step in a successful application development cycle. The next challenge is to be able to understand how the application is used by the customers. Do users hang up in one particular place? Are abandoning rates high for the type of the application being developed? Are some paths in the application not being used at all? Once statistics around application usage have been gathered, one can then take this back into the design phase, make the adjustments to the application flow, and deploy an updated version of the application. With the Lexee platform, all of this can be done in a matter of hours or days, not weeks.

Section 2 below provides a high-level overview of the Lexee architecture, whilst Section 3 mentions future directions for research and development. In Section 4 we summarize the key points of this paper.

2. High-level architecture

The high-level solution architecture is outlined in Figure 1. Once the application dialogue design has been finalized, developers can proceed and build the application (voice site) flow. This is done using the Lexee tool called CX Builder. CX Builder allows users to rapidly build the application flow, upload prompts, create recognition grammars and define interaction with customer exposed RESTful APIs using GET and POST protocols. An example of what the CX Builder interface looks like is shown in Figure 2.

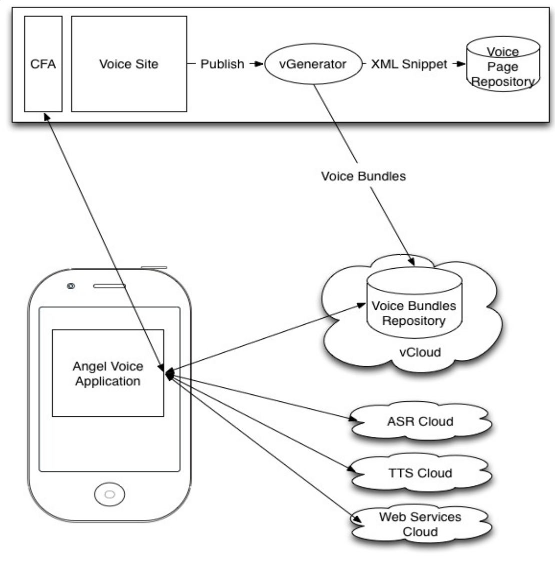


Figure 1: *Lexee high-level solution architecture.*

After the application flow has been built and tested using the CX Builder tool, users can generate and upload the VoiceBundle to the VoiceCloud (vCloud). The VoiceBundle is a package created by the Lexee tool called VoiceBundle Generator (vGenerator) which contains the following components: application flow formalized in SCXML [1], media files (prompts, images, and so on), grammars needed for successful interaction with an ASR resource formalized in GRXML [2], list of TTS prompts required for successful interaction with a TTS resource, configuration files defining any global application settings. The creation of such a voice bundle is fully automated within the Lexee framework, and results in a voice bundle being packaged together and uploaded to the voice cloud.

Further work on the mobile application is carried out using the SDK developed for a particular platform (iOS, Android, just to name a few). The uploaded VoiceBundle is fetched from the cloud and integrated into the mobile application. The SDK provides several components for application development: SCXML browser, ASR manager, TTS manager, media manager, web services manager and CX Analytics manager.

SCXML browser is used to interpret and execute the SCXML contained in the voice bundle, similar to a voice platform executing VXML [3] in IVR applications. ASR and TTS managers are components responsible for mediating between the SCXML browser and ASR and TTS resources, respectfully (whether embedded or in the cloud). Media manager serves the function of playing the sound files and displaying the images during the execution of the application flow, while web manager serves the function of mediating the interaction between the SCXML browser and any external resources. Finally, CX Analytics manager logs customer interaction with the application into the CX Analytics reporting system – this provides very useful information that can be used to determine the transactional success as well as customer mode of interaction (speech vs. touch), amongst many other things. Once the mobile application has been created with the help of SDK and has undergone all the necessary checks and approvals (e.g. Apple app store approval, PCI compliance [4]), it can be published to the application store and made available for purchase and download.

Any further updates and changes to the application once it has been submitted to the application store are extremely easy to carry out using the Lexee platform. The ease of making changes to the application and the speed of making these updates available to end users are a distinctive feature of the developed solution.

3. Future work

As mentioned in Section 1, mobile applications are becoming extremely widespread in our lives, and there is a growing expectation of the availability of more natural machine-human interfaces. The advent of Siri has shown how natural language processing is driving the future of mobile and telephone applications. It is planned that future work will encompass the addition of natural language understanding capability to the Lexee platform.

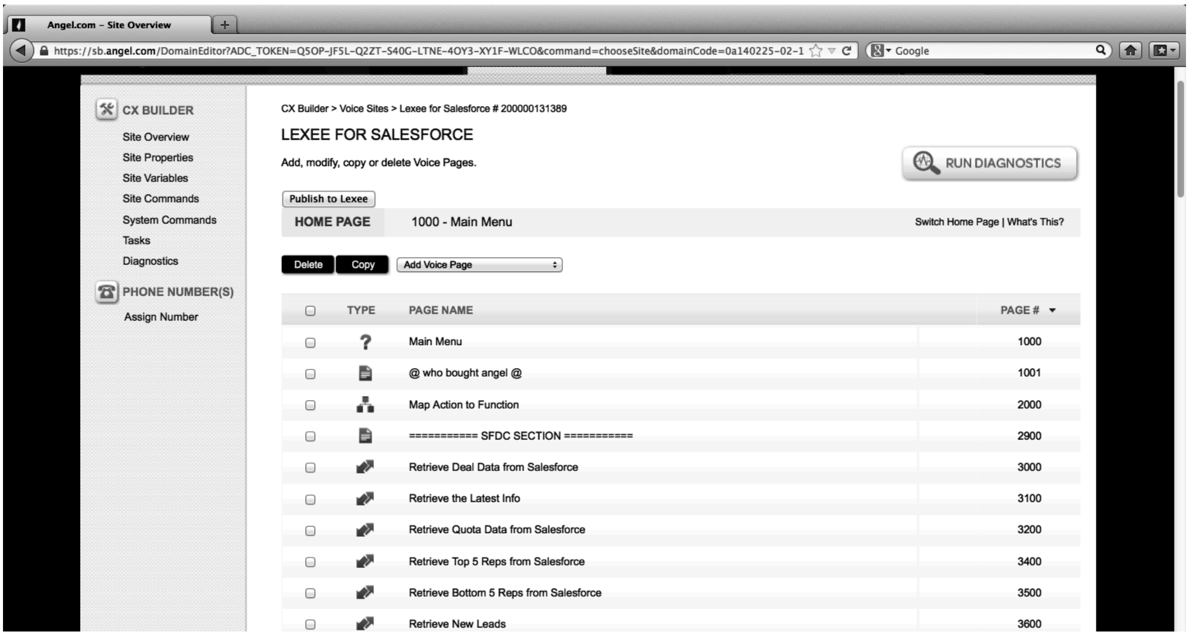
Another area of research will be focused around the use of embedded ASR and TTS technology in the mobile applications. Embedded speech technology has its advantages and disadvantages compared to its server-side implementation, and research is currently on-going to find optimal uses of embedded speech technology for the Lexee platform.

4. Conclusions

The paper presented a new cloud-based platform Lexee which enables developers to rapidly build and deploy mobile applications. Alongside the tools that come as part of the Lexee SDK, Lexee provides a framework designed to analyze application performance and make updates to design with minimal effort involved. Future work will concentrate round introducing a natural language capability as well as embedding some of the speech technology.

5. Acknowledgements

The authors would like thank Rajesh Ramchander and Brandon Phillips for their valuable comments on earlier versions of this paper.



CX Builder > Voice Sites > Lexee for Salesforce # 200000131389

LEXEE FOR SALESFORCE

Add, modify, copy or delete Voice Pages.

HOME PAGE 1000 - Main Menu

□	TYPE	PAGE NAME	PAGE # ▾
□	?	Main Menu	1000
□		@ who bought angel @	1001
□		Map Action to Function	2000
□		=========== SFDC SECTION ===========	2900
□		Retrieve Deal Data from Salesforce	3000
□		Retrieve the Latest Info	3100
□		Retrieve Quota Data from Salesforce	3200
□		Retrieve Top 5 Reps from Salesforce	3400
□		Retrieve Bottom 5 Reps from Salesforce	3500
□		Retrieve New Leads	3600

Figure 2: *Example screen of the Lexee tool CX Builder used to create the flow.*

6. References

[1] Barnett, J. et al., “State Chart XML (SCXML): State Machine Notation for Control Abstraction”, W3C Working Draft 6, December 2012; http://www.w3.org/TR/scxml.

[2] Hunt, A. and McGlashan, S., “Speech Recognition Grammar Specification Version 1.0”, W3C Recommendation, March 2004; http://www.w3.org/TR/speech-grammar.

[3] Oshry, M. et al., “Voice Extensible Markup Language (VoiceXML) 2.1”, W3C Recommendation, June 2007; http://www.w3.org/TR/voicexml21.

[4] https://www.pcisecuritystandards.org/security_standards.