

*Web-based visualization and corpus management for language modeling*

## Weekly Progress Report #11

### Customer

*Prof. Dr. Dietrich Klakow, Spoken Language Systems Group, Saarland University*

### Developer

Tiyash Basu, Bhavesh Bhansali, Supratim Das, Satabdi Ganguly, Clayton Greenberg, and  
Rohit Kumar Gupta

### Tutor

*Rafaella Antonyan*

### Submission Date

05/02/2016

#### Project description:

The high level goal of this project is to create a functional, web-based graphical user interface (GUI) for the language modelling toolkit developed by the Spoken Language Systems Group, Saarland University (LSVLM). When possible, the design of this GUI must be close to that of the existing non-functional GUI in order to provide a consistent user experience for existing users. The GUI must extend the functionality of the LSVLM toolkit and existing GUI in the areas of corpus management and visualization of provided outputs (Experimentation). The intended users of this GUI are expected to know the basics of language modelling, but not command-line tools.

## Progress

This is the second to last week of the final iteration. We are working on three main areas: completing the language model architecture implementation, generating concordances in the experiment app, and making use case videos for the benefit of the users. In our previous tutor meeting, we agreed that since the final product is hosted on the web, we do not need to prepare resources for installation and deployment as the product has already been deployed.

### what has been done

The interpolation models are mostly done. They render correctly, but some work remains on how to generate the ARPA files as this is not a trivial task when the vocabularies are not the same.

### ongoing work

Once the interpolation models are done, we expect the skip models will be easier. Creating a special box for a non-custom modified Kneser-Ney will be a bit more of a challenge, but as a last resort, we could just promote the already implemented “copy” feature to cover this requirement.

### plans for the next week

We have another meeting with the beta testers next Wednesday where we will get last minute feedback. After this meeting, we plan to meet with the customer briefly to prioritize their suggestions and go over any final meeting formalities.

## Problems

The authentication issue has progressed. The corpora are now stored in the cloud. The cloud service has been contacted in case they are able to open their authentication service to us. Otherwise, we could use the SFB DokuWiki accounts, the CoLi accounts, or a fully separate database.