

Web-based visualization and corpus management for language modeling

Weekly Progress Report #8

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

15/01/2016

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 nonfunctional 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 week was the final week of the second iteration. As such, we focused on polishing the documentation which is due on Friday. We had an internal meeting on Tuesday to distribute tasks for the deliverables and a customer meeting on Friday. At the customer meeting, we presented the new features for this week and all of the documentation. Since LDAP authentication has now been implemented, some users have already begun using the site. We discussed their feedback and experience. We are planning to have a user questionnaire and dedicated beta testers in the final iteration. This will help us prioritize may-have features.

what has been done

- LDAP authentication
- Updated Requirements
- All Use Cases
- Class Diagram
- All Sequence Diagrams
- All Paper Prototypes

ongoing work

- We want to create instruction videos out of the Paper Prototypes
- The language modelling design page is in progress but not complete
- Scrolling on the surprisal graph is in progress but not complete
- The concordance list still needs to be implemented
- Further user validations should be implemented
- Further visualizations would be good

plans for the next week

We plan work on the few remaining must have features and conduct the user questionnaire.

Problems

Using RZ for authentication, while originally desired by the customer, requires too many outstanding dependencies, so the customer is happy to use his own LDAP directory instead. If this becomes different in the future, we will change the code then.