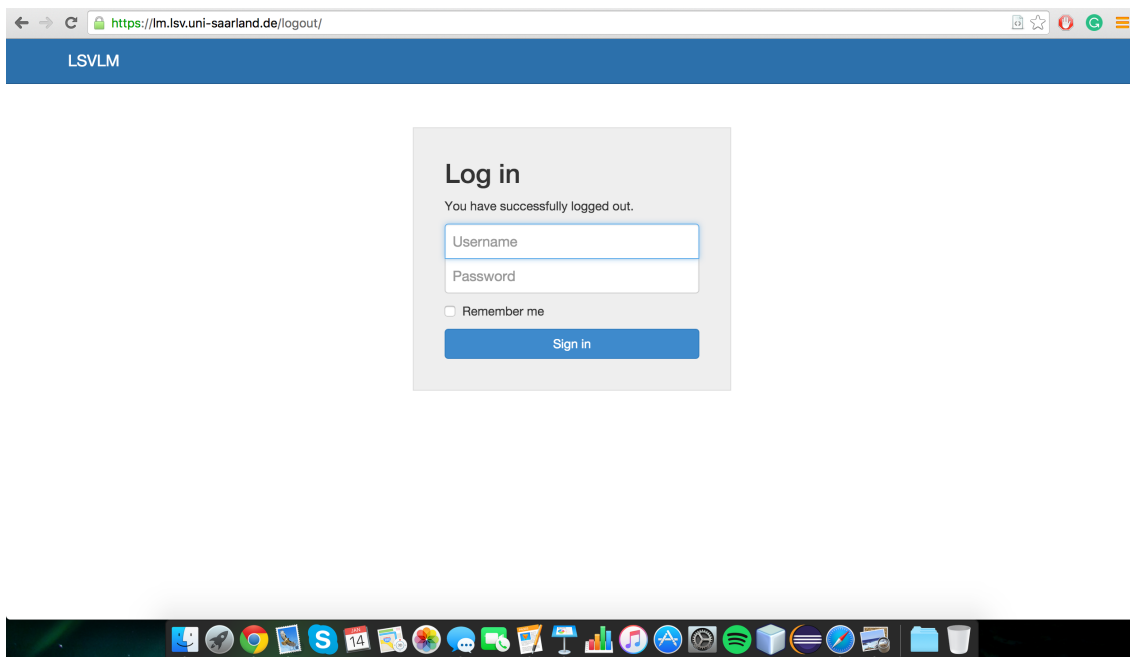
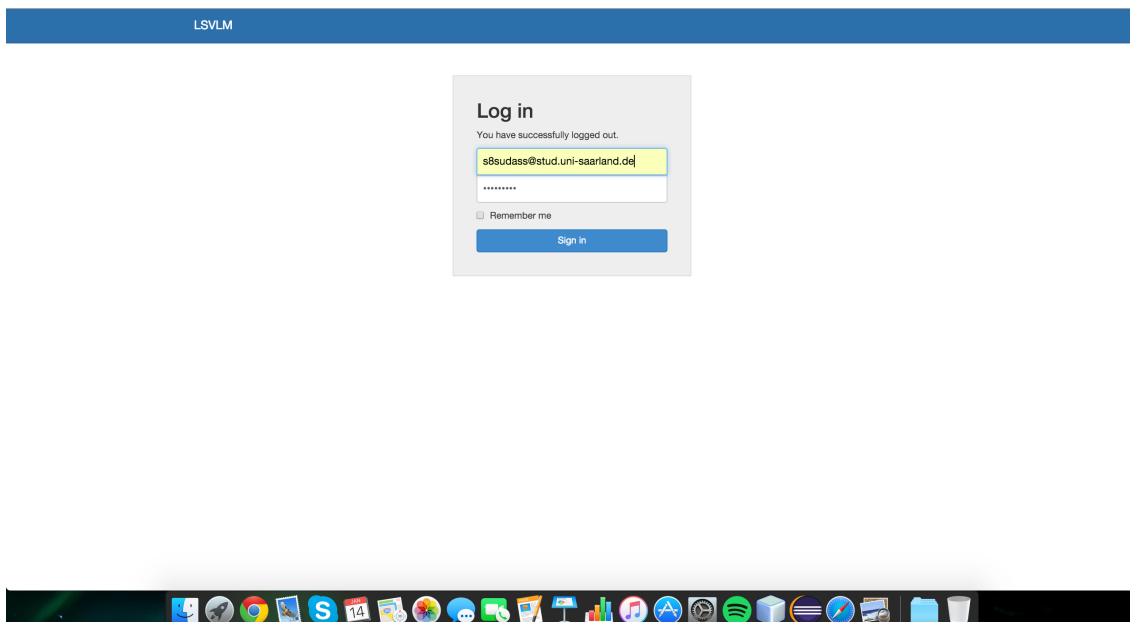


Paper Prototype: Corpus Management: Add new corpus

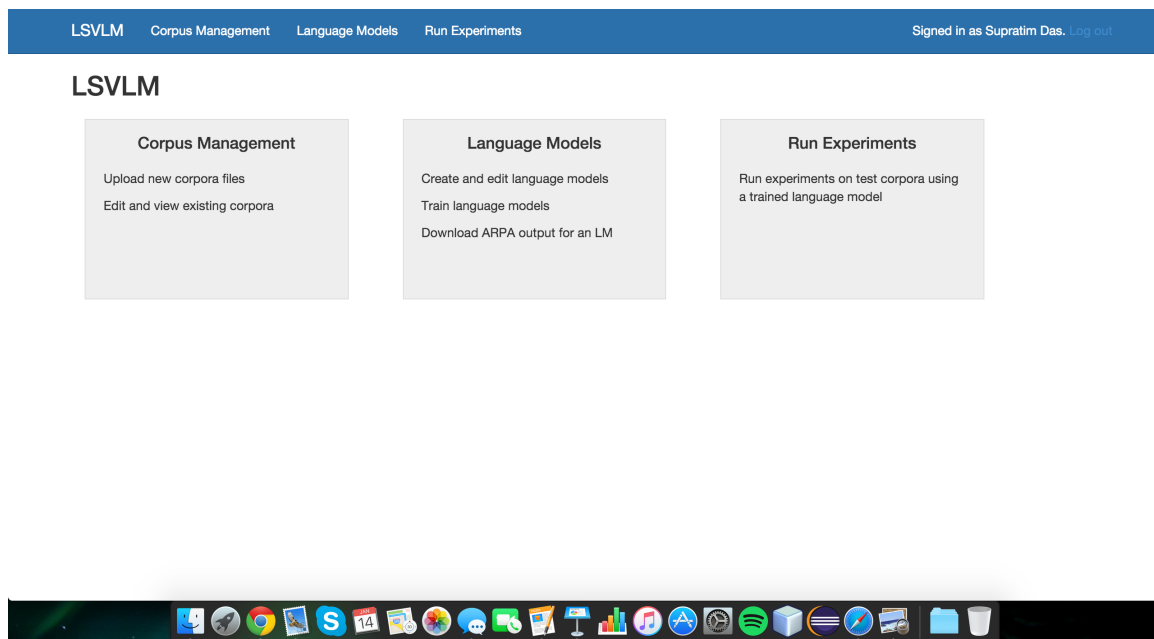
1. The researcher logs onto the 'LSVLM' GUI website.



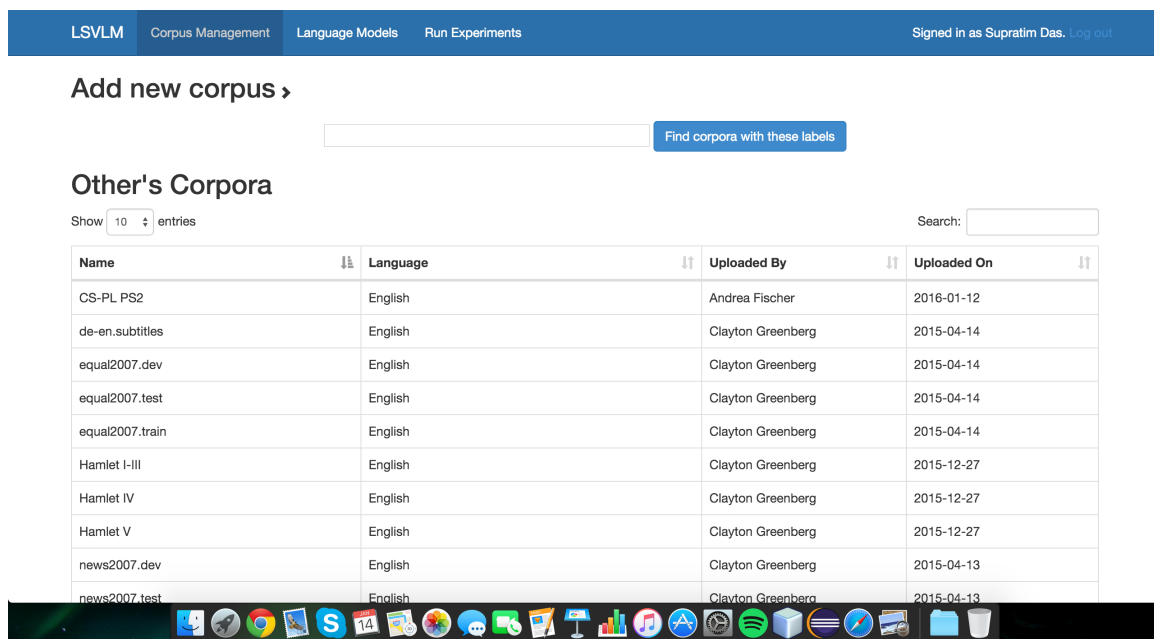
2. The researcher enters his username and password



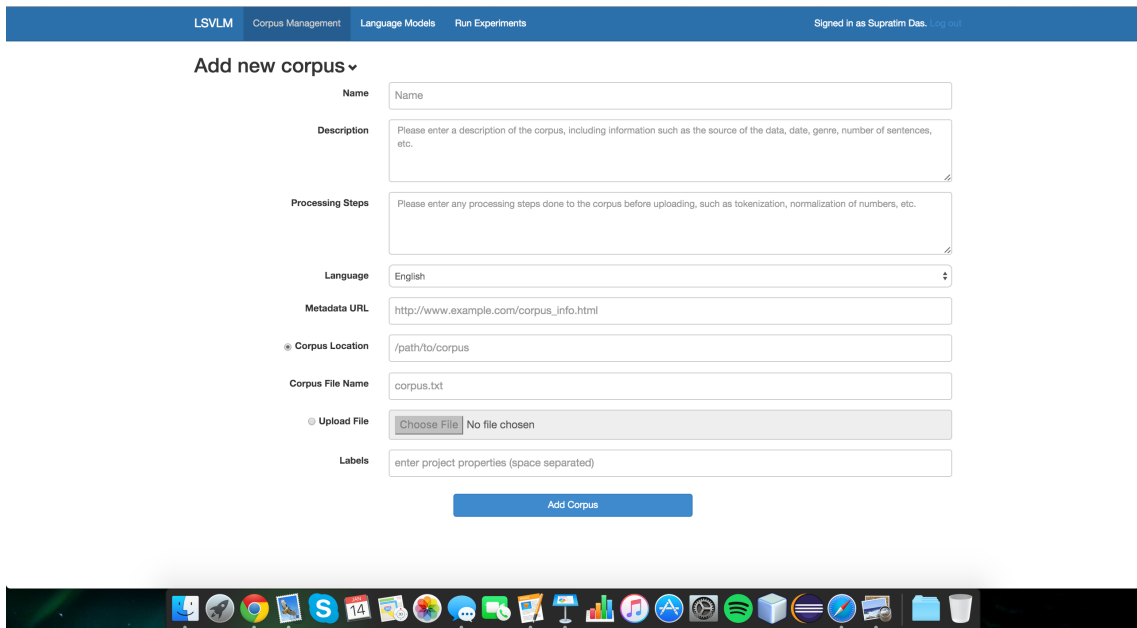
3. The system displays function groups, namely, Corpus Management, Language Models, Run Experiments.



4. The researcher selects “Corpus Management”.



5. The researcher selects “Add Corpora”.

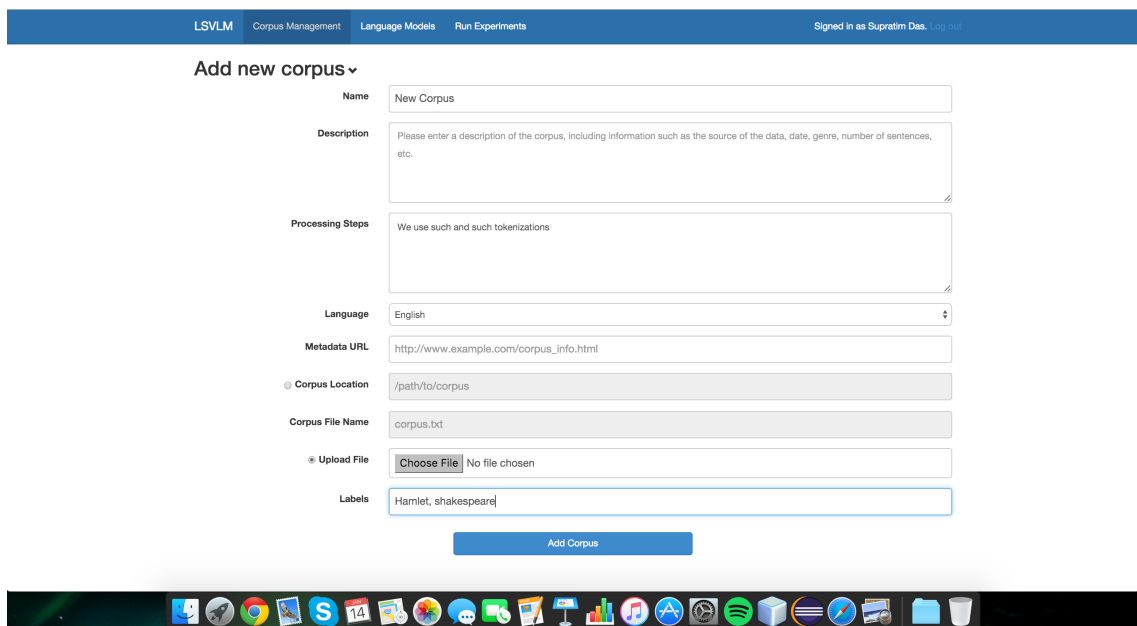


The screenshot shows the LSVLM Corpus Management interface. The top navigation bar includes 'LSVLM', 'Corpus Management', 'Language Models', and 'Run Experiments'. The user is signed in as 'Supratim Das'. The 'Add new corpus' form is displayed with the following fields:

- Name:** A text input field with the placeholder 'Name'.
- Description:** A text area with the placeholder 'Please enter a description of the corpus, including information such as the source of the data, date, genre, number of sentences, etc.'
- Processing Steps:** A text area with the placeholder 'Please enter any processing steps done to the corpus before uploading, such as tokenization, normalization of numbers, etc.'
- Language:** A dropdown menu currently set to 'English'.
- Metadata URL:** A text input field containing 'http://www.example.com/corpus_info.html'.
- Corpus Location:** A text input field containing '/path/to/corpus'.
- Corpus File Name:** A text input field containing 'corpus.txt'.
- Upload File:** A button labeled 'Choose File' next to the text 'No file chosen'.
- Labels:** A text input field with the placeholder 'enter project properties (space separated)'.

An 'Add Corpus' button is located at the bottom of the form.

6. The researcher fill in the necessary fields to add new corpus model.



The screenshot shows the LSVLM Corpus Management interface with the 'Add new corpus' form filled out. The fields are populated with the following information:

- Name:** 'New Corpus'
- Description:** (Empty text area)
- Processing Steps:** 'We use such and such tokenizations'
- Language:** 'English'
- Metadata URL:** 'http://www.example.com/corpus_info.html'
- Corpus Location:** '/path/to/corpus'
- Corpus File Name:** 'corpus.txt'
- Upload File:** 'Choose File' button next to 'No file chosen'
- Labels:** 'Hamlet, shakespeare'

The 'Add Corpus' button is visible at the bottom of the form.

7. The researcher fails to add new corpus because of the constraints.

The screenshot shows the 'Add new corpus' form in the LSVLM interface. The form has the following fields and errors:

- Name:** New Corpus
- Description:** Please enter a description of the corpus, including information such as the source of the data, date, genre, number of sentences, etc. (Error: Enter either description or metadata_url)
- Processing Steps:** We use such and such tokenizations
- Language:** English
- Metadata URL:** http://www.example.com/corpus_info.html (Error: Enter either description or metadata_url)
- Corpus Location:** /path/to/corpus
- Corpus File Name:** corpus.txt
- Upload File:** Choose File | No file chosen (Error: Please select a file.)
- Labels:** Hamlet, shakespeare

An 'Add Corpus' button is located at the bottom of the form.

8. The researcher finally fills in all the details. The following page appears.

The screenshot shows the 'Confirm corpus upload' page in the LSVLM interface. The page displays the following information:

- Confirm corpus upload**
- 520322 bytes
- 2207 lines, 9687 words, 520322 characters
- Is this content correct?
- Continue** button

9. The researcher finally fills in all the details. The following page appears: Confirmation of the upload.

LSVLM

Corpus Management

Language Models

Run Experiments

Signed in as Supratim Das. [log out](#)

View Corpus

Edit Corpus

Delete Corpus Record and Files

Delete Corpus Record Only

Name	New Corpus
Label	New Corpus English Hamlet, shakespeare
Language	English
Size	520322 bytes
Count	2207 lines, 9687 words, 520322 characters
Uploaded By	Supratim Das
Uploaded on	Jan. 14, 2016, 9:06 p.m.
File location	uploads/new-corpus.txt
Sample Text	
Metadata URL	
Description	This corpus is about...
Processing Steps	We use such and such tokenizations