

Progress Report

Andrew Sciotti, sciotti2

Which tasks have been completed:

The following tasks have been completed:

- Setup environment (for python 2.7)
- Debugged code so that it was working in original state
- Identified the location where to hook into the EductionalWeb system to apply auto-text summarization to the returned results
- Implemented two text summarization prototypes, extractive using shallow NLP & abstractive using deep NLP

Which tasks are pending:

- Implement summarizers into rest of the code base (currently sitting separate using an example text paragraph)
- Fully hook in the summarizer to display on the EductionalWeb browser (see “challenges”)
- Explore improvements to the shallow NLP extractive text summarization (stemming, stop word removal, word embedding, etc)

Are you facing any challenges:

- 1) Errors during running

I cannot get the full workflow to work. The scoring function inside the “get_explanation” function of model.py freezes during computation. See

<https://github.com/Asciotti/CourseProject/blob/master/model.py#L311>

However, this only occurs when I run it via the app.py, if I were to manually run the scorer, though ipython/command line, it has no issues. Because of this, I cannot fully close the loop.

- 2) The data (para_idx_data) is not representative of what exists on the publicly hosted website

The sourced “explanations” from the <http://timan102.cs.illinois.edu/explanation/> website seem to be derived from the full text of the textbook, but the data provided in the git repo appears to only be snippets/headers of the text.