

Technology Review Proposal

Review and Comparison with Similar Tools of the spaCy Toolkit

Text Information Systems - CS 410

Daniel Zurawski - dzuraws2@illinois.edu (Coordinator)

Vedprakash Mishra - vrn2@illinois.edu

Gassan Soukaev - soukaev2@illinois.edu

The topic of our technology review is to evaluate the ease of use and effectiveness of the spaCy text mining package and compare it with other text mining Python packages, such as Metapy, Sklearn and NLTK. The spaCy package attempts to distinguish itself from other Python text mining packages by providing a streamlined approach to mining text for developers. Therefore, we want to focus on evaluating how useful the toolkit is for intermediate developers who are familiar with essential text information methods.

To accomplish our evaluation, we plan to use the established packages (Sklearn, Metapy, NLTK), to mine information from a corpus of e-mails. Then we will attempt to mine the corpus using spaCy. Finally, we will evaluate the useability and flexibility of spaCy compared with the other packages. We also plan to provide examples of code to illustrate how coding using these packages looks like.