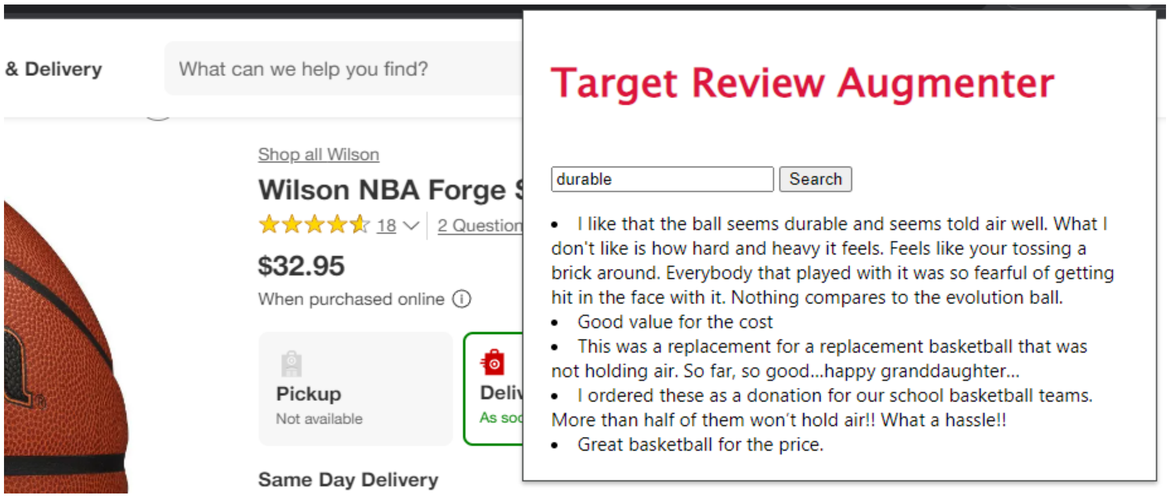
Team NLPro Project Progress Report

Members: Ethan Howes (edhowes2@illinois.edu), NetID: edhowes2

Progress So Far

A simplified version of the Target Review Augmenter extension has been completed. When the user is on a Target.com product page they can submit a query and the extension will use BM25 to search the reviews for that product and display results. The front end of the extension was implemented using HTML, CSS, and JavaScript. Some visual enhancements were made to the extension but the UI is still in progress. The extension's backend is in Python, so a runnable Python flask server was set up so that the Python code to do the BM25 search is callable from JavaScript. This flask server is manually run by the user with a simple Python command included in the readme. The BM25 search was initially implemented using Metapy. However, due to issues described in the Challenges/Issues section, the rank_bm25 Python package was used instead. Rank_bm25 fits the web app style of this project better than the functionality Metapy provides. Finally, the RedCircle API was integrated to fetch product reviews from a Target product page. The user will also need to set up a free RedCircle API account and use their API key to use the extension. These steps will also be clearly described in the final readme.

Hours Tracked:
10/30 - 1.5 hours (learning Chrome extensions and how to call Python code from it)
10/31 - 1.5 hours (learning Flask and how to call JavaScript from it)
11/4 - 2 hours (passing HTML input and current tab URL to Flask, displaying Flask response)
11/4 - 0.5 hours (CSS learning and tweaking)
11/5 - 1.5 hours (integrating Metapy scoring with RedCircle sample JSON)
11/7 - 1 hour (fix Metapy hanging when called from Flask - had to set threaded=False in Flask)
11/11 - 2 hours (writing setup readme, setting the project up on another machine to test)
11/12 - 4 hours (integrating RedCircle API and moving from Metapy to rank_bm25)



- I like that the ball seems durable and seems told air well. What I don't like is how hard and heavy it feels. Feels like your tossing a brick around. Everybody that played with it was so fearful of getting hit in the face with it. Nothing compares to the evolution ball.
- Good value for the cost
- This was a replacement for a replacement basketball that was not holding air. So far, so good...happy granddaughter...
- I ordered these as a donation for our school basketball teams. More than half of them won't hold air!! What a hassle!!
- Great basketball for the price.

Remaining Tasks

- Preprocessing reviews using the Natural Langauge Toolkit (NLTK)
    - Stemming
    - Stop word removal
- Tweak BM25 parameters for better performance
- Sentiment analysis with NLTK
    - Intelligently aggregate the sentiment scores of each review
    - Display comparison between star rating and sentiment score in extension UI
- Finalize Readme
    - Add API key process to the README
    - Describe the functionality of the extension and provide screenshots.
- Overhaul the UI for an intuitive user experience
    - Display a loading wheel when submitting the RedCircle API request

Challenges/Issues

The primary challenges faced thus far were with Metapy. Firstly the Meta toolkit documentation website (https://meta-toolkit.org/) was inaccessible during this first stage of development. This was due to some sort of privacy error with the site ("You cannot visit meta-toolkit.org right now because the website uses HSTS"). This made Metapy more difficult to work with.

Despite the Meta toolkit website not working, I implemented the extension BM25 search with it initially. This implementation was clumsy since I could not find a way for Metapy to query documents on the fly, the documents must be written to a file on disk that Metapy reads from. This meant I had to retrieve the reviews with RedCircle, write them to a file locally, and then query them with Metapy instead of passing the array of review strings from Recircle to a Metapy function. It worked, but it was a very error-prone and messy design.

After getting Metapy working initially I started to encounter this error: "Process finished with exit code -1073741819 (0xC0000005)" with seemingly random queries. For example, I would have a reviews dataset for reviews about a highlighter and when searching for the term "basketball" the server would exit with this error. I would expect the query "basketball" to return nothing from a dataset about highlighters instead of crashing. However with some queries like "and" or "color" this worked fine and scores were returned.

In the end, I decided that Metapy was not the correct tool to use for my needs in this project. I discovered another Python library called rank_bm25 that supports passing in an array of documents to run BM25 on and still allows the user to tweak the BM25 parameters. To replace the preprocessing Metapy offers, I will use NLTK for stemming and stop word removal.