Final Write Up
CS/INFO 4300: jdt98, jfp87, td334, msf239

A. Project Statement

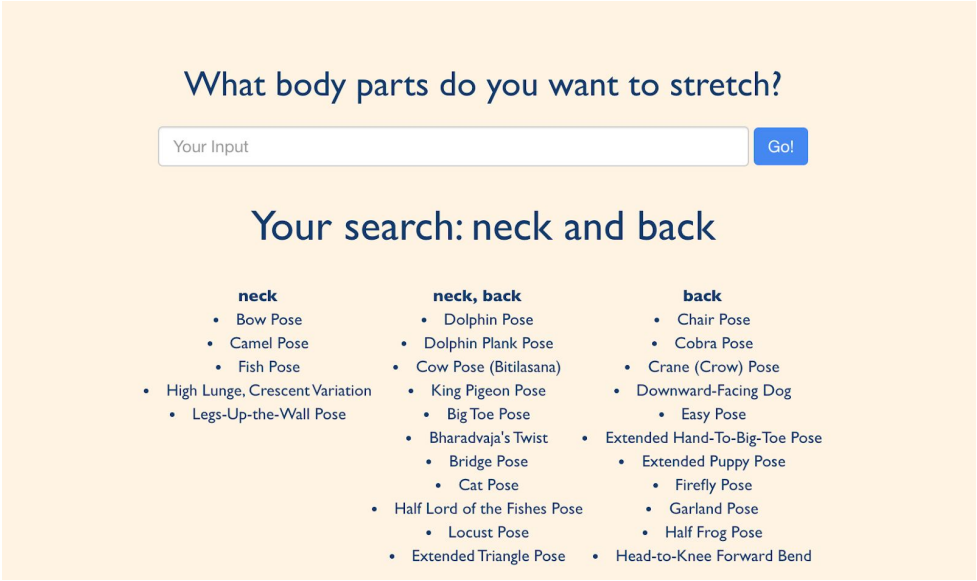
Goal of our application: To provide users with relief for body pains through stretches and exercises.

- **Input:** Area of body pain or soreness
- **Output:** 10 relevant poses + ML generated routine of <5 exercises
- **Use case:** During this time when all of us are spending more time sitting and staring at screens, our bodies are beginning to feel the effect of sedentary lifestyles. If the user has neck pains, the user would input “neck” into the application. The application would then output possible stretches, yoga positions, and exercises to the user that could be helpful to relieving their pain.
- **Techniques used**:
    - Cosine similarity: for ranking search results
    - Boolean search: for locating potentially relevant poses
    - Edit Distance: autofill implementation for misspelled search queries
    - Rocchio: feedback for vector space model: “ Are these results helpful?”
    - Sentiment Analysis: using Reddit discussions to rank popularity of poses
    - Machine Learning: to predict future searches and generate “routines”

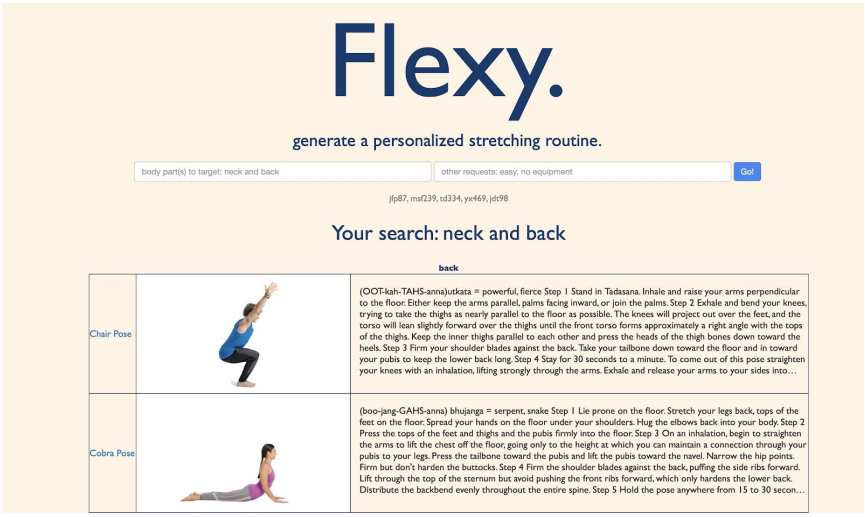
B. Iterations

Input: <u>neck and back</u>

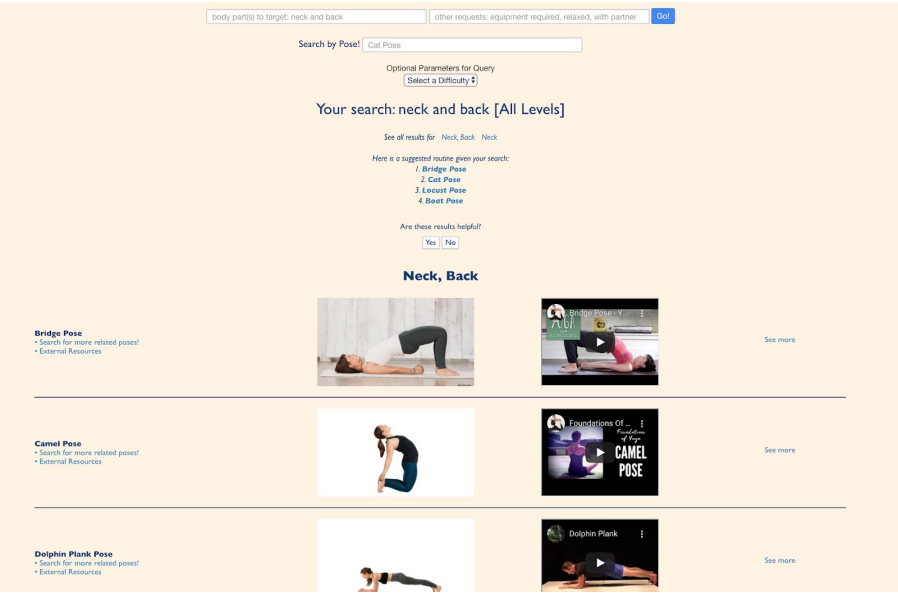
Prototype 1:



Prototype 2:



Final Prototype:



Improvements in UI/ Functionality:

a. Hyperlinks to jump to the top of each section.
b. Categories sorted by relevance
c. A suggested routine generated via ML
d. Added resources- image, youtube video, link to written instructions- for user to complete the pose accurately.

Improvements in Quality of Search Results:

1. For the first two prototypes, the relevant poses were listed in alphabetical order (obviously not ideal).
2. The final iteration has results from the boolean search displayed via rank using an adjusted cosine similarity measure.
3. Using this query as an example, the top ranked results are much more relevant than the top results on previous queries. As somebody with back & neck pains- Camel Pose and Plank Pose were both professionally recommended to me by physical therapists. (although it might seem counterintuitive to be doing planks but strengthening the core is helpful in improving posture and thus helping with back pain). Of course, every body is different, but this feedback gives us good insight that our search algorithms are appropriately tuned.

C. Summary of Work Distribution

Our work distribution was relatively well-balanced based on the schedule we set up and the tasks that we assigned each other. We had an idea of what each one of us had to do, and we executed the plan. Further, each of us contributed to the final writeup at the end of each sprint.

Jake Polacek (jfp87):

- **Major roles**: Deployment, Management
- Modified Flask template to properly display information retrieved from running a boolean search
- Updated query parsing implementation
- Implemented pose website linking
- Implemented edit distance to tell users if they input incorrect body parts
- Incorporated the Google Trends social component into the ranking system
- Synonym dictionary to allow us to include similar body parts in IR process
- Sorting of final body part categories by relevance to the original search
- Managed Heroku Deployment

Jonathan Tran (jdt98) :

- **Major Roles:** Backend, Data Scraping
- Data scraping (*The Ultimate Guide to Stretching and Flexibility,* Yoga Journal.com, htmls for directly related youtube videos and added on UI, Reddit- for popularity of poses)
- Data wrangling: converted yoga journal descriptions into step by step, intro and remarks sections
- Added UI autocomplete for pose search, display ML routine results
- Implemented Reddit popularity rankings and Rocchio
- Trained naive bayes classifier to rank difficulties for all the poses and stretches without difficulties
- Sped up backend by recomputing json loads and stores to improve search time from 8 seconds to 3 seconds

Matt Frucht (msf239):

- **Major Roles**: Frontend UI
- Responsible for interface of site- table design, buttons, card.html format

Teresa Datta (td334):

- **Major Roles:** Data, Search
- Ideation: conceived of chosen project idea
- Data Wrangling: Scraped, cleaned, manipulated Yoga journal dataset; conducted categorization algorithm.
- UI: small updates to improve aesthetics- incorporating links to different parts of the app, strategically limiting shown results.
- Machine Learning: Employed bootstrapping + Decision Trees to predict next search recommendation and unsupervised clustering algorithm to generate routines
- Search Algorithm: incorporated cosine similarity, rocchio, boolean search for both body part search bar and pose search bar

D. Acknowledgements of External Resources and Code Repositories

The Ultimate Guide to Stretching and Flexibility - Our dataset for stretching. We used images and words from this text.
Yoga Journal .com - Our dataset for yoga poses. We used images and words from this text.
https://www.youtube.com/watch?v=fVxIbQb8w58 - For an explanation on how to use datalists
https://towardsdatascience.com/scraping-reddit-data-1c0af3040768 - tutorial on how to scrape reddit with PRAW
https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html - external library used for TF-IDF and tokenizing, as well as bigrams and other features
https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html - external library for Multinomial Naive Bayes
https://jinja.palletsprojects.com/en/2.11.x/ - jinja documentation, to help me write jinja
https://www.w3schools.com/html/html_youtube.asp - documentation on how to embed video in html
https://www.w3schools.com/tags/tag_select.asp - documentation on how to use html select tag
https://www.tutorialspoint.com/flask/flask_sending_form_data_to_template.htm - flask tutorial on how to send data to back end and create a new view
https://www.crummy.com/software/BeautifulSoup/bs4/doc/ - library for scraping data

E. Discussion

Our app is different from similar apps (https://www.healthline.com/health/fitness-exercise/top-yoga-iphone-android-apps#simply-yoga, https://www.medbridgego.com) because not only do we provide images, step-by-step descriptions and videos, but we are also free (whereas similar products require in-app purchases or a $300 yearly subscription for physical therapy) . We include the ability to rank searches on an additional field that the user can specify based on cosine similarity, and allow for creation of simple routines. Further, we also allow searching by body part and pose, which is not something that is supported by competing apps. Finally, we allow for re-searching for similar poses or stretches if you find one that you like. So you can find similar items to a “crow pose” very easily. Overall, our product is unique because it offers additional functionality for users at no cost.

F. Ethical Concerns

We wish that our app be used properly by individuals in conjunction with the recommendations of their medical professionals. We are not physios and do not have medical training to provide aid to people who use our app with the endorsement of the medical community. We also acknowledge that the images and models used to demonstrate the poses are not representative of the diversity of the backgrounds and body types of our intended users.

Further, we wish that users acknowledge the Yoga Journal .com and *The Ultimate Guide to Stretching and Flexibility* . This is their work that we have scraped, and it is only right that the true authors get full credit for their hard work.