CMPSC 497 - Cognitive Computing

BehrendSync

Club Recommendation System for Students

Introduction

University clubs and organizations are the lifeblood of any great university. Clubs offer great activities for current student and residents of the university's surrounding area. They can provide valuable service to the community (e.g., cleaning up streets, volunteering at senior living homes, or cleaning up the beaches), endless hours of fun through activities (e.g., bingo nights, fashion shows, lip-sync battles), an atmosphere where ideas can thrive (e.g., professional speakers, support groups, and research nights). Clubs are undoubtedly one of the most important parts of the university ecosystem.

Penn State Behrend uses an online community management system called OrgSync. OrgSync provides an API to each university in order to supply them a platform with which to manage everything pertaining to clubs and organizations from registration to activity planning. When a student at Penn State Behrend goes through orientation, they are required to set-up their account for the Behrend version of OrgSync, BehrendSync, and can begin looking at all of the clubs available on-campus. If they are interested in a club they find, they can sign up directly on BehrendSync without having to go through traditional means of signing up. BehrendSync provides a one-stop shop for everything related to clubs and organizations in higher education institutions.

One of the running problems with BehrendSync is that the student has to find clubs manually based on what they are interested in. This means that unless they already have a club in mind, they will get back many clubs after searching; clubs that might or might not be relevant to what they are searching for. For example, when a student wants to search for art-related clubs on BehrendSync using the term "art," there are a few results. Among these results is the John M. Lilley Library department page and the Martial Arts Club. It can be said that these two results might not be the best possible results returned from the system.

Our system in its current form will work off of the current system, but will tailor the results to the student's interests while aiming to scrub out the results that would otherwise not pertain to these interests. The system will allow a student to write a small "about me" into the main part of the system or, if the student prefers to do less work and save time, import their Twitter profile

into our system. With one of these two bits of information, our system will analyze their text and use Natural Language Querying inside Watson Discovery in order to provide a detailed list of clubs that are closely related to the student's "about me."

On November 22, 2017, BehrendSync sent out a university-wide email that specified, in detail, that BehrendSync will be moving to a new platform that will provide "a fresh new look and upgraded features." This project is solely based on the current system implemented on the OrgSync API and we hope, as a group, that the new system will include a functionality such as what we have created.

System Implementation

Restrict query results in about me to top one for the club collection and top one for the constitution collection

Report with steps and screenshots and presentation slides:
a. Details of what you did. This could include tools, techniques, data sources, special processing requirements, problems and solutions, redoing experiments with more/different data/training/conversations/etc.

Our project was implemented in node.js, utilizing express and socket.io to host a webpage and process user input. This user input was then routed to the Watson Discovery Service as a Natural Language Query - the query fetches the title and the body of the most relevant document and returns them to our server, which then utilizes socket.io to serve the data to the user. Club constitutions and profiles were collected from OrgSync and entered into separate collections in the Watson Discovery Service - this was done to check results against each other, as clubs' mission statements and their OrgSync profiles contain a surprising amount of conflict.

We were forced to create our own database of clubs and constitutions due to being denied access to the OrgSync API. Therefore, we were forced to refine our data rather than having all of the data on hand through the API. Our results append the clubs constitutions one at a time.

Results Assessment

The query utilizing the “about me” statement ran as expected - Discovery efficiently processes the natural language query formed from the statement read in by the user, and returns the document with the most matching entities. Unfortunately, the natural language query regards each word as a unique entity and therefore can give poor recommendations when given short sentences, such as “I enjoy chocolate chip cookies” (this statement returns the Tone-Acious group’s constitution). Even restricting the output to only the top result from each collection of documents led to poor input data producing poor output data - however, this problem seems unavoidable when using the Watson Discovery Service without a great deal of data augmentation and protocols to reject or supplement poor data. This being said, our query works extremely well for even two to three sentences that describe an individual.

System Improvements

Acquiring the OrgSync API from the start would have made this process far more efficient. Being denied access from the OrgSync API after a lengthy wait process forced us to fall back to our back up plan. Having access to the API, instead of creating our own backup database of specific clubs, would have given us the ability to recommend potential activities that pair with clubs, as well as potentially creating a chatbot to instantly recommend clubs based off general keywords.

Improving the front-end to simplify and make it look cleaner would have been nice thing to add rather than a rudimentary structure that is currently in place. For example, connecting a database of images for the club recommendations would have been a nice touch to the front-end, but there was more of a focus on making sure the service was actually working. We wanted to focus on making sure that the club recommendation service was fully functioning before we focused on improving the front-end. Along with the front-end, linking Facebook would have given us more access to data such as Facebook likes rather than just a Twitter profile.

Will it succeed?

Basing our system off of how club search is performed in BehrendSync to begin with, we believe that our Club Recommendation System would be a more efficient way to search for clubs. Our Club Recommendation Service works great and would allow students to be exposed to clubs that they would not have been exposed to on the current system. Given that BehrendSync is switching platforms in the next couple of weeks, we hope that a similar function is provided in the new platform and will allow students to be exposed to more clubs through a similar functionality. Overall, our Club Recommendation feature is used to help students gain access to

as many clubs as they possibly can and we believe that this feature would help them achieve that goal.

References

[1] Orgsync.com. (2017). *Penn State Behrend*. [online] Available at: https://orgsync.com/browse_orgs/308 [Accessed 5 Dec. 2017].

[2] “Getting Started with the API.” *IBM Cloud*, IBM, console.bluemix.net/services/discovery/.