CS4241 Final Project README

Arthur Lockman, Isamu Nakagawa, Tucker Haydon

December 15th, 08:00 AM

Project Description

The goal of this project was to design and demo a new possible design for the O2CM website. O2CM is a website used to display the results of ballroom competitions.

As of December 2016, the website is in need of a update. While the website was functional, it was neither intuitive nor did it use spacing and styling well (pictures shown in the appendix). Links in the website are displayed in simple lists, results are shown in basic HTML tables, and page space is used inefficiently. Moreover, the results of the competitions are a pain to interpret. Instead of listing couples' and judges' names, the table lists numbers that then must be referenced at the bottom of the page to look up names.

We recognized two things:

1. O2CM was in need of a visual update

2. O2CM was a trove of data waiting to be interpreted/visualized

With this in mind, we formulated our objectives:

1. Improve website navigation

2. Improve results displays

3. Visualize data

Frameworks, Technologies, & Achievements

User Surveys

As a preliminary step of our design, we put together a survey of eight questions to present to users of the O2CM website. The goal of this was to use user feedback to gauge what the good/bad parts of the website are and to tailor our design for users. We posted the survey on a number of ballroom facebook pages and subreddits and received 86 responses. The survey questions are included in the appendix; results are zipped in the repository.

Overall, users were upset with the look of the website and the amount of navigation it took to find their desired results. Navigation was composed of a cumbersome steps involving ctrl-f, clicking, search boxes, and drop down boxes. When results were finally found, users had to scroll up and down, matching numbers in the results table to number associated with couples. Lastly, users were upset by poor lack of space webpage usage; results and lists are displayed on the left quarter of a page, and the other three-quarters were left blank.

On the other hand, users did like the tabular display of the results. Once a user had matched numbers, it was easy to understand how many points they got in each round and whether or not they had advanced.

With these results in mind, we formulated the first two objectives: cleaning up navigation and improving results displays.

Data Scraper

We wanted to to use and display real data, but did not have access to the O2CM dataset. To circumvent this, we built a webscraper that scraped data off of the O2CM website. The webscraper navigated to every round, for every dance, at every skill level, for every competition of the website; pulled parsed the results of the round off of the page; and saved it as a JSON object to be loaded into our database.

This webscraper was built in javascript and run with node. We used the Request library to make the HTTP calls and the Cheerio library to navigate and scrape data off of each page. Since the requests were done asynchronously, we used ES6 promises to manage the callbacks. Not all webpages were scraped, but an order-of-magnitude approximation for the number of webpages on the site is as follows:

$$10^3 \text{ competitions} * 10^1 \text{ skill levels} * 10^1 \text{ dances} * 10^1 \text{ rounds } = 10^6 \text{ webpages}$$

If you listened closely enough, you could head a small webserver screaming in the distance every time this scraper was run.

Firebase

Firebase was used to store all of the parsed data. We chose to use Firebase for two reasons: the service was free (good for college students); Firebase stores JSON, which was convenient since the scraped data was output as JSON.

Restful API

A RESTful API was created as an intermediary between the Firebase database and the frontend. The frontend website queried the RESTful API to get data regarding the list of competitions, skill levels, competitors, dance names, points, etc.

React

The front end of the website was designed using Facebook's React library. Custom React components were created for each of the different states the frontend

could be in (main page, competition results, etc), as well as for displaying the competition results and data visualizations.

Contributions

Arthur Lockman

1. Set up Firebase database

2. Designed entire REST API

Isamu Nakagawa

1. Designed frontend using React

2. Connected frontend & backend. Utilized REST API to retrieve data from the server and display it on the client.

Tucker Haydon

1. Wrote O2CM data scraper/parser

2. Designed data-displaying tables and charts

Lessons Learned

Dirty Data

One of the lessons we learned throughout this process is that real-world data is 'dirty'. When parsing data off of the O2CM website, it was incredibly frustrating to extract precisely the data we wanted. Much of the scraping involved matching regex or specific words or patters, but occasionally the data displayed on the website did not match the regular pattern, and the scraper would break.

For example, on almost every single results page, the number associated with a couple is of length 3 and the number associated with a judge is of length two. The scraped we initially built looked for numbers of length 2/3 and associated them with judges/couples. On a few pages in the harvard competitions, a couple would occasionally have a number of length 1, breaking the rule. The result was that the number and name were both incorrectly parsed.

Situations like this occurred throughout the process and it was frustrating to have to go back and modify the parsing rules every time an exception was discovered.
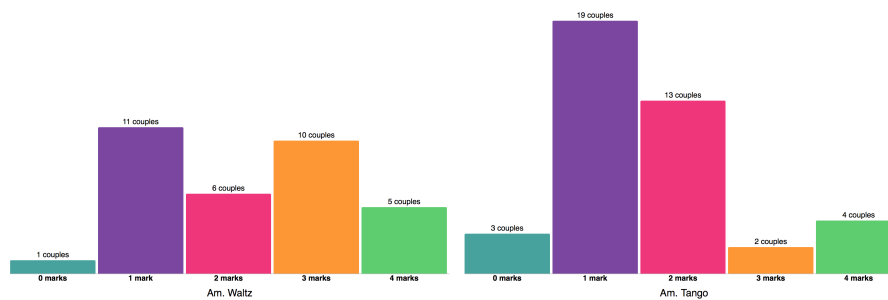
Appendix

O2CM Photos

All O²CM Competition Results

2016

Metropolitan DanceSport Championship	December 4th	1790 hits
California State DanceSport Championships (NQE)	December 3rd	1223 hits
Big Apple Dancesport Challenge	December 3rd	1496 hits
Starlight Dancesport Championship	December 3rd	1425 hits
2016 DSB Collegiate Challenge	November 19th	3846 hits
Chinook DanceSport Classic	November 19th	346 hits
Spartan Invitational Ball	November 19th	1374 hits
Tufts University Ballroom Competition	November 13th	2519 hits
SnowBall Classic, Hyatt Regency Vancouver BC	November 12th	2132 hits
Triangle Open DanceSport Championships	November 12th	655 hits
Claremont Intercollegiate Showdown	November 12th	426 hits
Liberty Ball DanceSport Challenge	November 6th	979 hits
Badger Ballroom Dancesport Classic	November 5th	1212 hits
Brown Ballroom Dance Competition	October 30th	2745 hits
Quest For The Best	October 29th	581 hits
Devil DanceSport Invitational	October 29th	1078 hits
Chicago DanceSport Challenge (NQE) and WDSF Open	October 28th	2003 hits
Purdue Ballroom Classic	October 22nd	2079 hits
2016 UConn Husky Classic	October 22nd	1572 hits
Harvard Beginners 2016	October 16th	1462 hits
Cornell DanceSport Classic	October 15th	810 hits
Rocky Mountain DanceSport Grand Prix	October 14th	1148 hits
American DanceSport Festival	October 8th	2420 hits
25th Annual Yale Ballroom Dance Competition	October 8th	1931 hits
2016 Princeton Ballroom Competition	October 8th	2467 hits
Illini Dancesport Invitational	October 7th	1513 hits
Delta Cup DanceSport Gala	October 7th	526 hits
Carolina Fall Classic (NQE)	September 30th	3373 hits
CANADIAN OPEN DANCESPORT CHAMPIONSHIPS	September 24th	1715 hits
Wild Rose Ball	September 24th	516 hits
Defi Danse Mascouche	September 17th	1489 hits
Kansas City Dance Classic	September 10th	1015 hits
West Coast AAU Junior Olympic Games	July 22nd	667 hits
Gumbo DanceSport Championship (NQE)	June 24th	3095 hits
Summer Sizzler - NJ DanceSport Classic NQE	June 11th	3731 hits
Toronto Grand Cup	June 4th	1369 hits
Nebraska Invitational Dance Festival	June 3rd	1132 hits
24th Annual Calgary Open Dance Competition	May 14th	1475 hits
New England DanceSport Championship (NQE)	May 13th	2937 hits
Brooklyn Star Ball	May 8th	1079 hits
Island Fantasy Ball 2016	May 7th	2353 hits
Whistle Stop DanceSport Challenge	April 24th	577 hits
Stanford Cardinal Classic 2016	April 23rd	2225 hits
Charlotte Dancesport Challenge	April 16th	1373 hits
Irish DanceSport Gala	April 16th	2067 hits
2016 MIT Open	April 16th	11157 hits
Rhode Island College BDC	April 10th	1984 hits
Defi Danse Trois Rivieres	April 9th	1874 hits
Pacifica Amateur Ball 2016	April 9th	1144 hits
RPI Dancesport Competition 2016	April 2nd	1784 hits
USA DANCE National DanceSport Championships	April 1st	7480 hits
BAM Jam	March 26th	1745 hits
Rutgers DanceSport Competition	March 26th	2100 hits
Keystone Dancesport Classic	March 19th	1731 hits
Phoenix Dancesport Challenge	March 19th	1034 hits

Tufts University Ballroom Competition 2016
Final Results:
Provided by O²CM

Division	Age	Skill	Style	Competitor	
all	all	all	all	---OR---	OK

This competition is large and displaying the entire list is slow.
Please select from above filters to display some, or all, of the entry list then click OK

Adult Beeaux Smooth (WT)
1) 126 Jonathan Pee & Marilyn Sun - MA
2) 233 Nate Eckerson & Breanne Messier - MA
3) 269 Jay Grabon & Amanda Blaska
4) 179 Tucker Jaenicke & Julia Broni - MA
5) 235 Sean Lee & Annie Chen
6) 269 Salt Snyder & Emma Hodgdon - MA
7) 141 Graham Bright & Mia Kazman
8) 116 Patrick Reilly & Lucy Billings
---
9) 146 Ronmachai Jaroenset & Vivian Zhong - MA
10) 144 Robert LaFrance & Eva Huang
10) 300 Ian Dechene & Kaeya Katugam
12) 148 Theodore-Sauyet & Kylie Whalen
12) 246 Christopher Barbeau & Janel Guo - MA
12) 249 Saarthak Gupta & Hannah Blakely - RI
13) 279 Jeffrey Galczynski & Angela Ding
---
16) 119 PJ LeBlanc & Caroline desJardins-Park - MD
16) 157 Tucker Haydon & Christine Deidrich - MA
16) 163 Vance Catapang & Emily Fernandes
16) 255 Steven Chick & Francesca Fonzetti
20) 137 Sreyashi Sharmin & Noelle Austin - RI
21) 129 Izzui Ikram & Emily Small - MA
21) 190 Alex Schneiderman & Brittany Wendzel
21) 193 Ariel Sanchez & Gloris Wu - CT
21) 254 James Michielini & Aruni Kacker
21) 283 William Tait & Katherine Bell
26) 226 Jackie Chen & Dorothy Xiao
---
27) 107 Jeffrey Lin & Kayla Anderson - MA
27) 109 Michael Vollmer & Brianna Boyce
27) 113 Uku-Kaspar Uustalu & Nicole Lonergar - MA
27) 178 Jongn Wang & Samm Katcher
27) 196 Nino Salazar & Alana Richert - MA
27) 206 Aidan Pryor & Deshauna Hammond
33) 142 Elise Westerveit & Lauren Varanese - NJ
33) 175 Richard Lin & Briana Choi
35) 202 Hector Hernandez & Cathy Xue - CT
35) 272 Michael Chaoui & Mao-Chu O'Connell - MA
37) 156 Brian Naito & Sarah Guilong - MA
37) 170 Charles Weinacht & Bernadette Berman - RI
37) 212 Russell Azzarello & Mallory Stone-Murphy - RI
37) 226 Ben Dunbar & Sarah Sherman
41) 106 Stephanie Miller & Ashley Smith - NJ
41) 161 Mathew Ennis & Maraya Clark
41) 197 Jesus Bastardo & TBA TBA - MA
41) 220 Sanjia Sterlin & TBA TBA - MA

Adult Beeaux Am. Foxtrot (F)
1) 233 Nate Eckerson & Breanne Messier - MA
2) 179 Tucker Jaenicke & Julia Broni - MA
3) 126 Jonathan Fee & Marilyn Sun - MA
4) 269 Jay Grabon & Amanda Blaska
5) 146 Ronmachai Jaroenset & Vivian Zhong - MA
6) 144 Robert LaFrance & Eva Huang
7) 235 Sean Lee & Annie Chen
---
8) 119 PJ LeBlanc & Caroline desJardins-Park - MD
8) 137 Sreyashi Sharmin & Noelle Austin - RI
8) 255 Steven Chick & Francesca Fonzetti
8) 300 Ian Dechene & Kaeya Katugam
12) 269 Salt Snyder & Emma Hodgdon - MA
13) 157 Tucker Haydon & Christine Deidrich - MA
---
14) 116 Patrick Reilly & Lucy Billings
14) 129 Izzui Ikram & Emily Small - MA
14) 163 Vance Catapang & Emily Fernandes
14) 190 Alex Schneiderman & Brittany Wendzel
14) 193 Ariel Sanchez & Gloris Wu - CT
14) 226 Jackie Chen & Dorothy Xiao

New Website Photos

Am. Waltz: 0 marks — 1 couples; 1 mark — 11 couples; 2 marks — 6 couples; 3 marks — 10 couples; 4 marks — 5 couples

Am. Tango: 0 marks — 3 couples; 1 mark — 19 couples; 2 marks — 13 couples; 3 marks — 2 couples; 4 marks — 4 couples

Survey Questions

1. How do you use O2CM? Select all that apply.

   (a) Look up results for self
   (b) Look up results for another
   (c) Look up results for team
   (d) Look up results for an unknown competitor
   (e) Other (Please specify)

2. On a scale of 1-5, how easy is it to find your desired results?

   (a) 1 - Very Easy / Uncomplicated
   (b) 2
   (c) 3 - Reasonable
   (d) 4

    (e) 5 - Very Difficult / Complicated

3. How would you *initially* prefer to filter results? Select up to three.

    (a) By competition

    (b) By name

    (c) By style

    (d) By skill (newcomer, bronze, etc)

    (e) By age

    (f) By division (amateur/pro)

    (g) Other (Please specify)

4. On a scale of 1-5, how easy is it to recognize/understand the results?

    (a) 1 - Can immediately recognize/understand the results

    (b) 2

    (c) 3 - Can reasonably recognize/understand results

    (d) 4

    (e) 5 - Have difficulty recognizing/understanding the results

5. Do you have any suggestions for better displaying the results?

6. On a scale of 1-3, what is your interest in tracking results over time?

    (a) 1 - Very interested

    (b) 2 - Moderatly interested

    (c) 3 - Not interested

7. What other information would you like included on the website?

8. Do you have any other suggestions or comments about the website?