<p><b>What if we had a weekly BCS ranking for the OMPA?</b> <br>

If you have any appreciation for the BCS system in college football or an NFL mock draft, you'll appreciate the 'mock OMPA' python script that I rigged up a few weeks ago. For anyone that isn’t a swimming fanatic that grew up in Contra Costa County (i.e. most people), I’ll give some background: the OMPA is a highly competitive swim league in Orinda and Moraga, California. It’s comprised of 9 teams and 2000+ swimmers from a total population of less than 40,000 people. This tiny hotbed of swimming is remarkable for having helped produce over 17 Olympic swimmers. This includes Matt Biondi and quite a few other gold medalists (<a href="http://www.ompaswim.com/About\_OMPA.html">OMPA History</a>).. The biggest part of the OMPA league is without question the OMPA championship swim meet.</p>

<p> I’ve been connected to the OMPA since I was 6 years old. I swam in the league as a swimmer, coached summers in college, and have continued to coach for what is now almost 8 years. The OMPA has always been a super bowl like event for many of the families in the area, and mine was and is no exception. Ever since the league started posting swim times to an online database almost 15 years ago, it’s been easy to speculate what surprises might happen on the big day of the championship meet. </p>

<p>

However, we have always had to wait for the actual day of competition to see how the OMPA teams stack up against each other. In the past, we had to wait for each dual meet between two teams to be swam. We would wait for the times to be uploaded from all of the different dual meets that weekend. Then and only then could we, as coaches, look at rankings and estimate by gut which particular teams seemed to look pretty fast. Scoring a future OMPA championship after each and every week of dual meets would take 10s of hours each and every week. Scoring it by hand would also mean that you would have to pick each swimmer’s best 3 events of the 5 events they likely have in the database. Picking relays would be even more tedious. </p>

<p> <b> Python to the rescue</b><br>

Thankfully, this is the perfect task for a computer program. More specifically, it’s the perfect task for python scripts. Using <a href="http://doc.scrapy.org/en/latest/intro/overview.html"'>scrapy</a>, a framework for building web crawling and data scraping ‘bots’, I was able to quickly obtain and sort all of the swimmer times from the <a href="http://www.crgwebservices.com/OMPA/">OMPA database website</a> into nicely organized json files. With another script, I wrote algorithms that figure out each swimmers 3 best events and the best relay combinations for each team. The script then scores the entire swim meet in a near instant (that’s 82 events of scoring the top 20 swimmers & 9 relays from a pool of 2000+ swimmers!). The script then outputs a nicely formatted meet results file of what an actual OMPA championship swim meet would look like as if it were swam at that exact point in time. The print out also has team scores broken down by age group, gender, stroke, or stroke within the age group. It is possible now to see which teams have the top 13up butterfly or the top 8under breaststroke rather than being limited to the traditional age groups. </p>

<p><b>Tracking changes throughout the season</b><br>

This means that rather than simply having to wait for the OMPA in August, we can now run the mock ompa at certain times throughout the season and track the changes from week to week. It means we can finally quantify just how much Sleepy Hollow out scores their seed position heading into the real OMPA (they always do). It means we can quantify which teams improve the most throughout the season. I have a few hypothesis that I’d like to test. For example, despite the end of season huge gap between the fastest and slowest teams, might all the teams start the season fairly close together? Perhaps the best teams simply have slightly more improvement week to week that compounds? The mock OMPA model will allow me to test this.</p>

<p><b>Running a mock Lafayette-OMPA</b><br>

Any non-OMPA team, who’s times can be web-scraped in a similar fashion, can be loaded into the mock OMPA and scored as if they were OMPA member teams. Because I’ve coached teams in both the OMPA and in Lafayette, I’ve always wanted to know exactly how these neighboring towns stacked up against each other (the county meet really isn’t a fair comparison). Springbrook usually beats all of the OMPA teams at County, but are they really that much faster? By loading their times into the mock OMPA, we can finally quantify that Springbrook is a solid 4th place OMPA team, and Rancho Colorados is somewhere between a 5th or 7th place team.</p>

<p><b>Predicting 2014 from 2013</b><br>

Each age group gender stroke (ie 9/10 girls freestyle) is handled by a separate web crawler, and loaded into it’s own JSON file for the Mock OMPA script to later handle. By setting the scraper to load the 10 to 11 girls freestyle into the JSON that usually contains the 11/12 girls freestyle, the mock OMPA is effectively run with everyone aged up to where they would be next season. This gives us a Mock OMPA for 2014 even when it’s still 2013.</p>

<p><b>Dual Meets</b><br>

Through testing my program, I realized that by not including a team name in my scoring function, I could effectively score the meet without them ever being loaded into the events. Eliminate all but two teams, and you have a mock dual meet. There was already a separate variable for scoring the relay events versus the individual event, so I just added another scoring scheme that resembles the much fewer dual meet points.</p>

<p><b>Some predictions based on the 2014 output</b><br>

As you can see from the output files below, I do not think this will be the year that Sleepy Hollow overtakes OCC. It’s been 10 years+ of OCC, and while Sleepy Hollow does seem to be gaining, the prediction remains OCC for now. I do think it’s possible that OPP or MCC make a run at 3rd place, as Meadow’s predicted lead over 4th and 5th place seems a lot smaller than what I remember past years actual results being.</p>

<p><b>The Code</b><br>

<a href="https://github.com/NickStefan/MockOMPA">Mock OMPA on GitHub</a></p>

<p><b>Example Outputs</b>

<br><a href="/static/mockompa2013.txt">2013 Mock OMPA using swim times from August 2013</a>

<br><a href="/static/mockompa2014.txt">2014 Mock OMPA using swim times from August 2013</a>

<br><a href="/static/mocklafomparcsblt.txt">2013 Mock Lafayette-OMPA</a></p>