

CSC7062 Final Assessment Report

Conor Bradley • 40108536 • cbradley75@qub.ac.uk

Database creation and design:

Dataset: The UFC fight statistics dataset chosen can be found here:

<https://www.kaggle.com/theman90210/ufc-fight-dataset>

The database was normalized as far as was appropriate for the intended functionality, chiefly the aim being to reduce replication and redundancy of data. The tables with the bulk of the data are the fights, fighters, eventsTitle and eventsVenue tables. See the below ER diagram in figure 1 to see the general relationships used for the database.

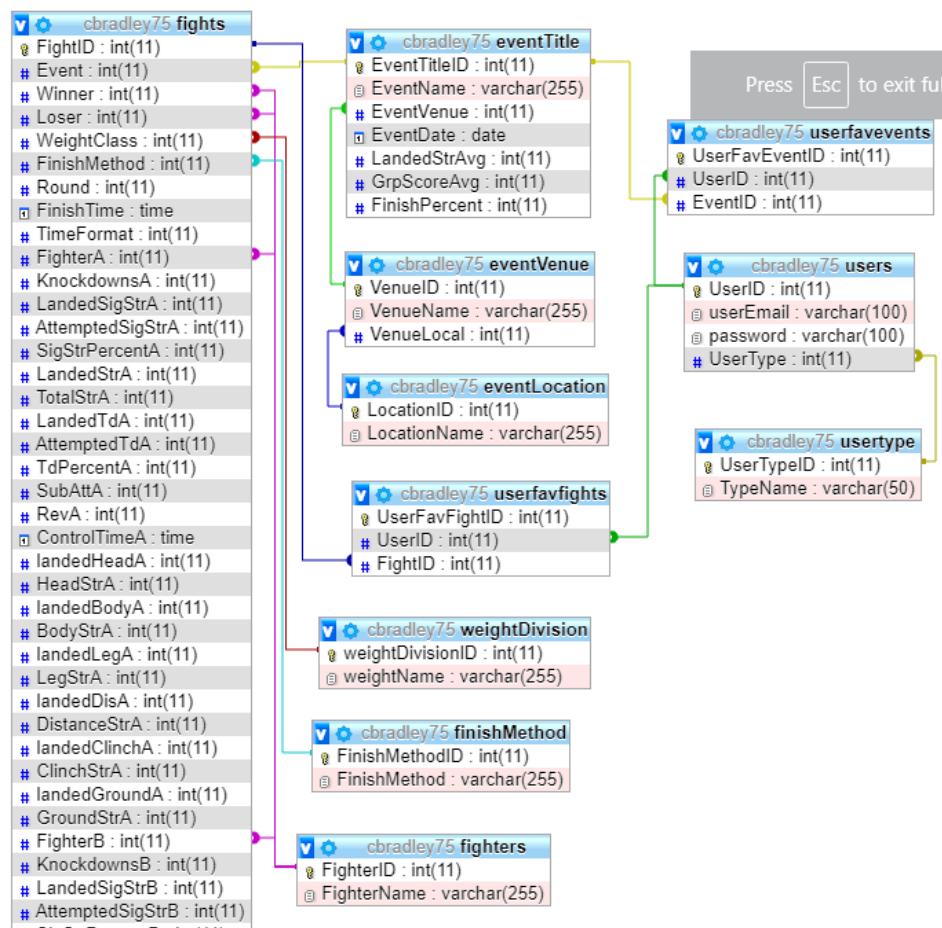


Figure 1: ER diagram of UFCFightData database.

The most major challenge to overcome with the chosen dataset was the lack of an accompanying date for a fight. This being necessary for filtering of the data and for informative purposes in general. A solution was implemented by obtaining a separate CSV file, filled with all UFC events including their date, from the UFC events wikipedia page's 'Past events' table[1]. This is also where the venue and location was obtained. Using a third party website the table's contents was outputted into a csv [2] referred to as the events csv. An attempt to parse this data using the wikipedia open API proved futile. These records were matched up to the existing datasets records by comparing the title fields, being the only point of reference to compare from, although there was

frequently slight variation between the two fields, e.g. In the dataset csv: *UFC Fight Night: DosAnjos vs. Ferguson*, was named in the events csv as *UFC on FOX Fight Night: DosAnjos vs. Ferguson*. And so the PHP `similar_text` function[3] was used within a while loop to cycle through both csv files at once, first holding the first value of the events csv and then cycling through every value of the dataset csv, if the similarity of the title column value string was above 95%, the row (including its title) from the dataset csv and the date, venue and location from the events csv was saved and written to a separate csv (named with the postfix of *_trim*) which in turn was further preprocessed and used to populate the database. Of course this resulted in not all event titles being picked up on the first run of this process. The 95% similarity criteria was necessary at first as some event titles were quite similar but different such as *UFC Japan 2* and *UFC Japan 3*. However see the `runOnceTrimEvents.php` file within the php folder for how the csv's were then trimmed down, removing the values already matched successfully. This can then allow for another pass through the similarity function of the remaining values, but at a lower comparison % criteria. The hope being that with each pass, the remaining events to get matched with decrease and with a few executions of this process, all events and so fights data would be successfully matched inputted. However due to time constraints, only two passes were performed, this still being sufficient to match and input nearly 4500 fights with foreign keys successfully.

Using PHP scripts the database tables were populated with the appropriate data. Besides the previous preprocessing explained above, many values from the CSV had to undergo additional preprocessing before input. A number of columns within the dataset had to be split into two separate fields, these columns being the columns for the strikes landed vs thrown by body target (strikes to the head, body, leg) and from position (strikes thrown vs landed from distance, clinch and ground) for each fighter of a fight. These had to be split into two separate columns, for example head strikes from fighter A originally being for example '35 of 56', split into two separate fields 'landedHeadA' and 'HeadStrA'. This being essential for processing and outputting the stats data later.

Formatting was also performed as part of the preprocessing of the data before input. Percentage signs had to be stripped from percent field values. Time based fields such as fight finish time had to be reformatted to concatenate an additional "00:" onto the value to match the SQL datatype format used. See *runOnceFights.php* for the code to perform this.

SQL prepared statements were used extensively in the populating of the database. Necessary as certain fields such as fighters name contained characters special characters such as ' in "Sean O'Malley" for example caused an SQL syntax error.

URL(s) to REST Application Programmable Interface (API):

GET: <http://cbradley75.lampt.eecs.qub.ac.uk/ufcFightStats/API/getStatsapi.php>

Events information and filtering:

Retrieve all events data:

Filter by date:

Filter by Average Landed Strikes per fight of event:

Filter by Grappling Score of event:

Filter by Finish rate (% of fights to end by stoppage):

Retrieve specific event's data:

Parameters:

?allevent

?allevent&date={newest / oldest}

?allevent&landedStr={highest / lowest}

?allevent&grpScr={highest / lowest}

?allevent&finRate={highest / lowest}

?eventid={#}

Fights information and filtering

Retrieve all fights data:

Retrieve all fights data from specific event:

Retrieve specific fight's data:

Retrieve specific fighter's fights data:

Filter by data:

Filter by Weight class:

Filter by Significant Strikes landed:

Filter by Grappling Score of fight:

?fights

?fights&eventid={#}

?fights&fightid={#}

?fights&fighter={fighters id}

?fights&date={newest / oldest}

?fights&weight={user select weight class}

?fights&SigStrScr={highest / lowest}

?fights&GrdScr={highest / lowest}

User Information:

Retrieve users favorited events:

Retrieve users favorited fights:

?favEvent&userID={#}

?favFight&userID={#}

Search Function:

Search for fights:

Search for events:

?searchfights={input search string}

?searchevents={input search string}

POST: <http://cbradley75.lampt.eecs.qub.ac.uk/ufcFightStats/API/postStatsapi.php>

Favoriting/Unfavoriting event:

Favorite:

Unfavorite:

?favEvent

?unfavEvent

Favoriting/Unfavoriting fight:

Favorite:

Unfavorite:

?favFight

?unfavFight

URL(s) to the software system that uses the above API:

- **Video Demo:** <https://www.youtube.com/watch?v=IDBnwzYjYd0>
- Landing page: <http://cbradley75.lampt.eeecs.qub.ac.uk/ufcFightStats/index.php>

The landing page has three key features. An upcoming events section, a most recent event breakdown preview section, and a search function section. The upcoming event section is currently static and would require manual change at present from the developer. Ideally this could be populated automatically by web scraping the UFC's own next events page [4] as an example resource, however due to time constraints, this was seen to be out of scope for this project and a stretch functionality for future development. The search function acts in a rudimentary fashion, consisting of a jQuery ajax[5] api call on page load, passing the inputted string as a GET parameter to the api which uses the `similar_text` php function previously mentioned to compare the string to firstly all fighters names in the fights table and return a JSON result of fights and appending them to the fights table of the `search.php` page, a separate ajax api call is then performing the same process except this time for the eventTitle table comparing to the title of the event. This search bar can also be accessed and utilised by the user through-out the site in the navbar.

- Events index page:
<http://cbradley75.lampt.eeecs.qub.ac.uk/ufcFightStats/events.php?page=1&date=newest>
- Fights index page:
<http://cbradley75.lampt.eeecs.qub.ac.uk/ufcFightStats/fights.php?page=1&date=newest>

The events and fights pages allow the user to filter by various fields, such as date, set to newest by default, grappling score (which is a calculated weighted average of different aspects of the fight related to ground game, such as submission attempts being weighted more heavily, takedown attempts and landed takedowns, ground strikes etc) and event multiple fields at once, specifically by adding a weight class selection on the fights page.

- Event page: <http://cbradley75.lampt.eeecs.qub.ac.uk/ufcFightStats/event.php?eventid=1>
- Fight page: <http://cbradley75.lampt.eeecs.qub.ac.uk/ufcFightStats/event.php?eventid=1>

Chart.js[6] was used to produce the charts of the site, and is clearly displayed on the event and fight breakdown pages. The coloring scheme for the fight page charts reflects the fighters relative rank for that particular fight, blue being lower ranked, red being higher ranked, notice the circle beside the fighters name indicating this [6]. The user may also favorite an event or fight by clicking the star button [7] provided they are logged in. This performs an ajax POST to the `favFight.php` or `favEvent.php` page which performs POST call to the api to add the event as a record of the `userfavfights` or `userfavevents` table respectively. And the same process occurs to remove them from the table if clicked again using the `unfavFight.php` and `unfavEvent.php` pages.

- Signup page: <http://cbradley75.lampt.eeecs.qub.ac.uk/ufcFightStats/signup.php>
- Login page: <http://cbradley75.lampt.eeecs.qub.ac.uk/ufcFightStats/security/login.php>

Upon sign up a user record will be created using SQL prepared statements to protect against SQL injections. The `signupprocess.php` file that is POST to from the `signup.php` page also performs a check of the database to ensure that email doesn't already exist and redirects the user back with a warning message and a prompt to login if it does. The bulma framework also provides a useful format check in these email fields for both login and signup pages[8]. On login a SESSION variable is created and set, unlocking other features of the site such as the favorite button, and the

userprofile page or the admin page. Upon logging out, the *logout.php* page will send the user back to the page from which they came, provided that page does not require a SESSION to be active [9]

- User profile: <http://cbradley75.lampt.eecs.qub.ac.uk/ufcFightStats/userprofile.php>
 - Login first: Email: john@ufc.com Password: ringo123
- Admin page: <http://cbradley75.lampt.eecs.qub.ac.uk/ufcFightStats/admin.php>
 - Login first: Email: bigDug12@con.com Password: password123

Links for these pages will appear on the nav bar depending on the user type that is logged in. It must be noted that in this application, a user type of public user can be created from the signup page, and an admin must be created directly by the developer in the database. The userprofile displays all the users' favorited events and fights in separate tables, with links to each. The admin holds two forms which may be selected to input new fights or events into the database. At present these forms are not active, however this would be a functionality to be implemented with future development.

Comparison of proposed and finished design and stretch functionality:

Most of the proposed functionality has been achieved to a large degree. One of the main features proposed being the ability for the user to filter events and fights by such things as date, strikes landed (e.g. for boxing fans), grappling score (e.g. for jiu-jitsu fans) or weight class. This however could be further expanded, one of the most obvious being to filter by year, a simple addition however time constraints forced prioritising of functionality. Another key proposed feature achieved was the display of stats in charts, again this can be further developed, the labels of the chart only show the red corner to reference the shades of color, this needs further development to include the blue corner. As mentioned previously the admin forms to add fights and events also need to be developed, these will be more complicated of a process as a lot of information, being the foreign key values must first be retrieved from the database to populate the dropdown options of the forms. Additional forms would also need to be added for the addition of new fighters, venues and locations etc to the database. Other functionality to be implemented, previously mentioned, being the landing pages upcoming event section being auto populated. A major addition that needs attending to is the addition of an API authentication key. At current a user can alter the databases fav tables by directly accessing the api. This is a security flaw, however time constraints leave this to be a stretch functionality to be further developed. Another key addition could be the addition of a fighters index, and individual fighter's stats page, which would run similarly to event/fight index and breakdown pages. Fighters could then be filtered to find those with the best grappling score, head/body/leg strikes landed, even best defense by calculating the number of strikes thrown by the other fighter of their fights and taking away those that landed. The data afford a myriad of possible interesting stats to be calculated and displayed.

REFS:

- [1]https://en.wikipedia.org/wiki/List_of_UFC_events#Past_events
- [2]<https://wikitable2csv.ggor.de/>
- [3]<https://www.php.net/manual/en/function.similar-text.php>
- [4]<https://www.ufc.com/events>
- [5]jQuery v3.6.0 <https://code.jquery.com/>
- [6]Chart.js v3.2.1 <https://www.chartjs.org/>
- [7] FontAwesome v4.7.0 <https://fontawesome.com/v4.7.0/>
- [8] Bulma v0.9.2 <https://bulma.io/documentation/form/input/>
- [9] Back after login : <https://stackoverflow.com/questions/5285031/back-to-previous-page-with-header-location-in-php>