Release Note: Web Crawler/Scrapper using Laravel6.0

The application is to crawl and scrap the info. Form the webpages. I am using Laravel6 Framework with 2 packages for Crawling the web pages **(laravel/dusk) dusk spyder** and the other for scrapping **(weidner/goutte)** the information from the webpages.

The database table schema is designed based on the reference website (http://www.mycorporateinfo.com/) provided in the email. I have used the normalization techniques to store the information of the organizations present the in the specified website.

Please use the following github link to clone the application into the project directory.

git clone git@github.com:kmdmustaq/laravel-crawler-scrapper.git

Please run the following commands to bootstarp the application

1. Add all the composer dependencies using the following command.

    composer install

    Thic compose will add all the dependencies along with laravel/dusk and widner/goutte

2. Once the composer install is finished its execution execute the following command so that laravel/dusk which will generate scaffolding in the application.

    php artisan dusk:install

3. Please setup your mysql database by providing them in **.env** file like
    DB_DATABASE=xxxxxx
    DB_USERNAME=xxxxxx
    DB_PASSWORD=xxxxxx

The configuration's for the added composer dependency's are already add in the cloned application.

4. Run Migration command to create database tables with the seeds (Initial Locations) tables in the database.

    php artisan migrate --seed

Starting the application:

Run the following command to bootstrap the application

    php artisan serve

The crawller is added as the test so please run the the test as follows.

<path-to-project>/vendor/bin/phpunit <path-to-project>/tests/Browser/duskSpiderTest.php

The above command will initially creates the fresh database by droping all the existing columns, crawls the whole website and collects the **web urls, titles, http_status_code, crawled status** and stores into the **crawlers** table.

Please have patience as the test command will take time to crawl the whole website. Or you can use the sql file provided in the application root path for the **crawlers** table.

After the crawling is completed please execute the following url in the browser to perform the scrapping.

http://localhost:8000/scrap

after completion of scrapping you will see a simple json response on the screen as follows

```
{
    data: {
        Total_urls: xxx,
        Scrapped_urls: xxx,
        messgae: "Success"
    }
}
```

with status code 200.

Thank you!!!