Project Proposal:

Indexing and retrieving Metadata from an Instagram Dataset

Jordi Morera Trujillo
Eric Presas Valga
Adrià Carbó Duch

1.- Project Overview and Goals:

The propose of this project is to build a web application that retrieves some data from an Instagram images dataset given by the teacher. Some new techniques to develop a web application will be introduced in our work in order to learn how this new tools work.

Inspired by the new kind of web applications that has been appeared around the last few years, which are using the work presented in 2009 by Ryan Dahl called Node.js, and it's support for Mongo DataBase.

The main goal of this project, as a deep back-end talking, consists on indexing the metadata from our images dataset to a Non relational database (MongoDB). Since this kind of databases has appeared, a lot of applications are taking use of it because of it's JSON formats for the collections. Since our metadata set is provided in JSON format, we are taking use of this kind of databases that store information in collections.

As a support for giving some "real" application to our work, we also propose to make a web application in order to simulate the main basic thing that you can do with a metadata dataset of images. Basically we are going to focus on GET and POST methods from REST applications.

2.- Project Background:

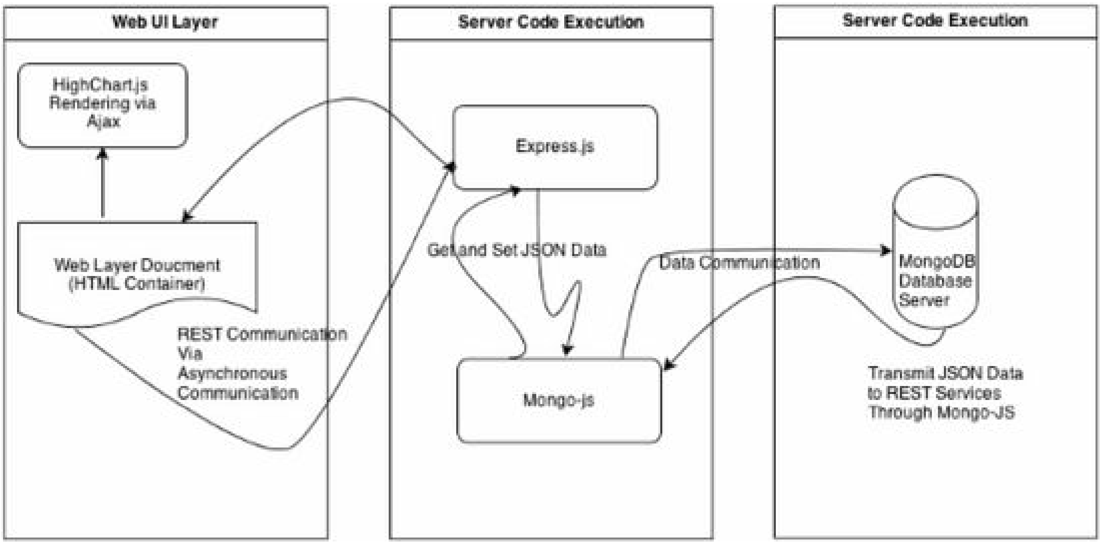
In this section, the architecture of the system will be introduced in order to provide some information about which kind of tools and frameworks that we are going to use.

The server side is structured in two main parts, the server part of our web application and the database server to provide the information.

- Application server-side: Is where all the information is processed, there we have the post and get methods that allows us to render the html pages. Also, all the information about the connections, frameworks used, public folders, and more are specified in app.js.

- Database: Is the server where the database is installed and working, usually it is placed in the same machine pointing at some other port or in the cloud since MongoDB offers free services for all the people that want to try it from cloud.

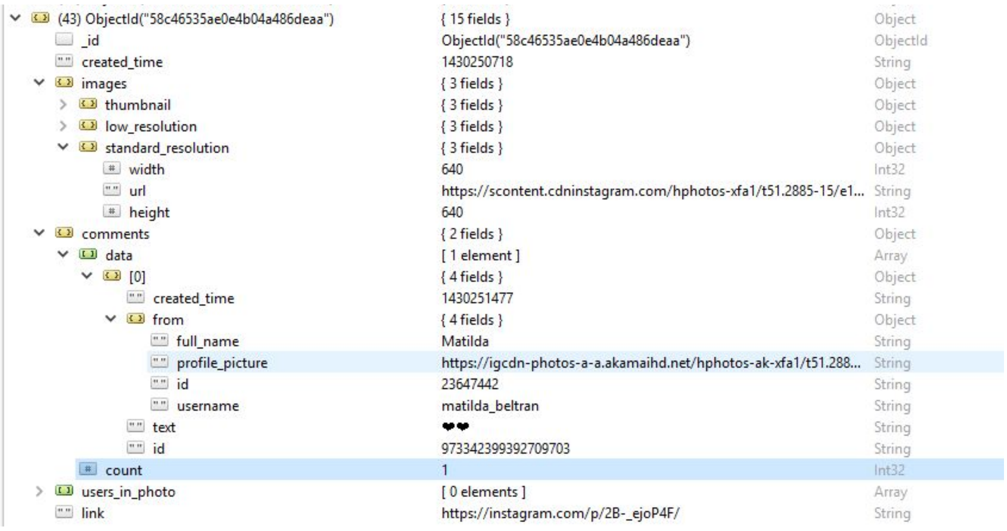
For the view side a HTML template engine called JADE will be used in order to render the dynamic HTML content.


Web UI Layer
Server Code Execution
Server Code Execution
HighChart.js Rendering via Ajax
Web Layer Doucment (HTML Container)
REST Communication Via Asynchronous Communication
Express.js
Get and Set JSON Data
Mongo-js
Data Communication
MongoDB Database Server
Transmit JSON Data to REST Services Through Mongo-JS

Project Requirements and Specifications:

MongoDB and JSON's structure

As said before, MongoDB is a NoSQL document-oriented database program that uses JSON-like documents. In our case we have a metadata set of images taken from instagram which structure looks as follows:



```
v (43) ObjectId("58c46535ae0e4b04a486deaa")   { 15 fields }                                                  Object
      _id                                     ObjectId("58c46535ae0e4b04a486deaa")                            ObjectId
      created_time                            1430250718                                                     String
  v   images                                  { 3 fields }                                                   Object
    >   thumbnail                             { 3 fields }                                                   Object
    >   low_resolution                        { 3 fields }                                                   Object
    v   standard_resolution                   { 3 fields }                                                   Object
          width                               640                                                            Int32
          url                                 https://scontent.cdninstagram.com/hphotos-xfa1/t51.2885-15/e1...   String
          height                              640                                                            Int32
  v   comments                                { 2 fields }                                                   Object
    v   data                                  [ 1 element ]                                                  Array
      v   [0]                                 { 4 fields }                                                   Object
            created_time                      1430251477                                                     String
        v   from                              { 4 fields }                                                   Object
              full_name                       Matilda                                                        String
              profile_picture                 https://igcdn-photos-a-a.akamaihd.net/hphotos-ak-xfa1/t51.288...   String
              id                              23647442                                                       String
              username                        matilda_beltran                                                String
            text                              ❤❤                                                             String
            id                                973342399392709703                                             String
        count                                 1                                                              Int32
  >   users_in_photo                          [ 0 elements ]                                                 Array
      link                                    https://instagram.com/p/2B-_ejoP4F/                            String
```

As we can see in the capture, every object contain the metadata that we will use in order to make queries and search images by different fields.

Node Js and Express

Node Js is a JavaScript runtime built on Chrome's V8 JavaScript engine. It uses an event-driven, non-blocking I/O model that makes it lightweight and efficient. The Node.js package ecosystem, npm, is the largest ecosystem of open source libraries of the world. As an asynchronous event-driven JavaScript runtime, Node is designed to build scalable network applications.

To design the web application, Express will be used. It is a minimal and flexible Node.js web application framework that provides a robust set of features for web and mobile applications.

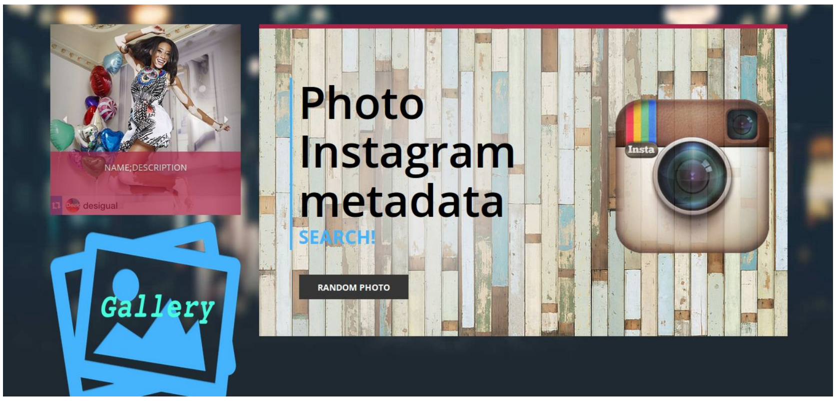
Web application

The web application will be developed with html templates, is an event-oriented language for developing and deploying the platform. It has its own programming language that exhibits a seamlessly integrated application server and event-oriented database management system, which consists on JavaScript functions.

The functionalities of the web application will be:

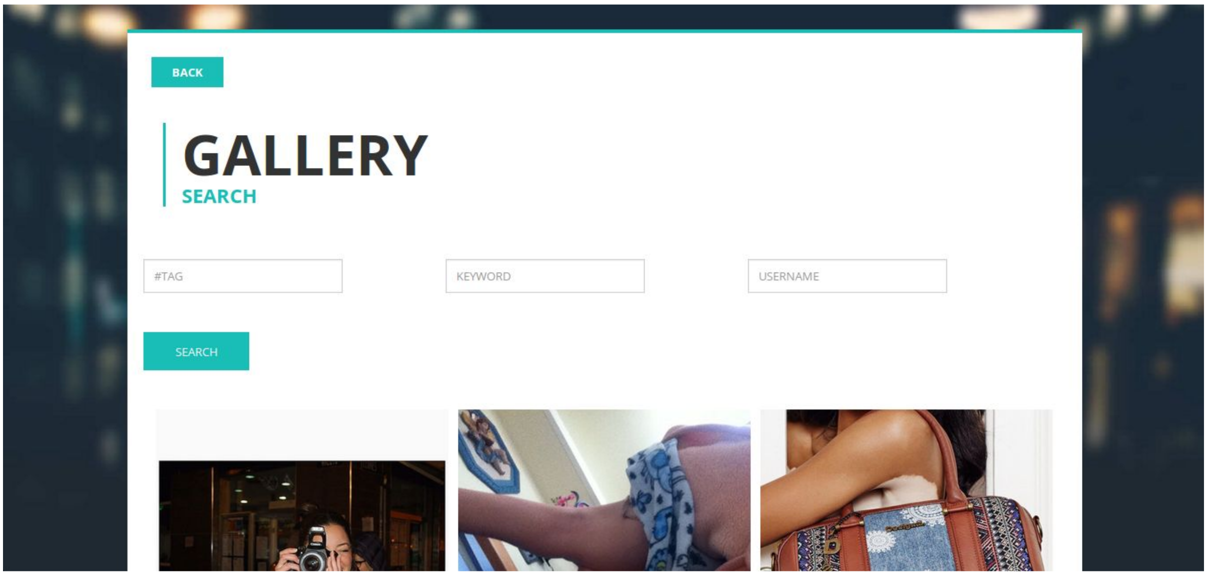
- Home page with a slider showing random photos.
- ‘Random’ page, accessed from home page. It will show a random photo and it’s data (name, comments, likes…).
- ‘Gallery’ page, accessed from the home page, it will request to the user to fill some fields of search (by tag, by fullname, by username), and it has to show all the images of the data base that has coincidence with any of the fields.

Home page



Slider on the top-left of the screen, *Random* link, and the *Gallery* link.

Gallery


BACK

GALLERY

SEARCH

#TAG

KEYWORD

USERNAME

SEARCH

Three form fields to fill (filters for photos), and a back button.