Migo Interview Assignment - read me

Project: Partner Discovery via Image Processing

Conceptualization:

Chose this project because it seemed the most complex and also it was something I wanted to do for a side project, Image processing. Spent the first 30 - 45 minutes coming up with the different modules needed to complete the task. Coding is always the easy part, so design was what I initially spent time on. Another reason to take this project up was, I initially toyed with the idea of completing the first two projects, because they involved searching, and based on the text entered into the text box I would either do an image search or search for results.

Approach:

The approach I decided to take was to start with a basic dynamic project in Java, a simple web page with a text box which takes in the input string and figure out if we can get an image out of it and if we did, parse the raw text out to the output.

Class structure was, one servlet class to handle the web requests, one validator class which would valid a URL, lastly an Image Processing class which would basically get the image from the web and then parse the text out of it.

Hypothetically: I would have another class which on the failure of not being able to parse and image, would do a Google search on the string entered and pass the results to the same text area which displays the raw text from parsing the image.

The validator class, used the regex to validate a url, and if it is only then would the url be passed onto the image processing module. I also figured if we could validate the image address by looking for extensions, but that requires a little more string manipulation, so if the string could not be resolved to a image I just pass the result onto a error page.

The Image processing class, used to process the image. Two parts to this, in the first part I just pull the image from the web and save it as temp file in the current working directory. And then next part I feed it into the web service which parses the image for text and returns a string.

Pitfalls faced:

My initial idea was to use the Tesseract.jar which would parse the image for text which would passed to it. But this was as rabbit hole I should not have gone into in hindsight. I spent considerable amount of time figuring how to install it and its dependencies, I felt like I had hit a wall. So took a break and figured out that I don't necessarily have to setup tesseract on my machine, all I have to do is parse the image for text. It flashed to me to look for web services online which I would do this for me for free. And there I found it. http://www.ocrwebservice.com/restservices

After that it kinda became apparent, how the flow would go further on. Only if I had figured this out instead of figuring tesseract out, I would have completed the other search project as well.

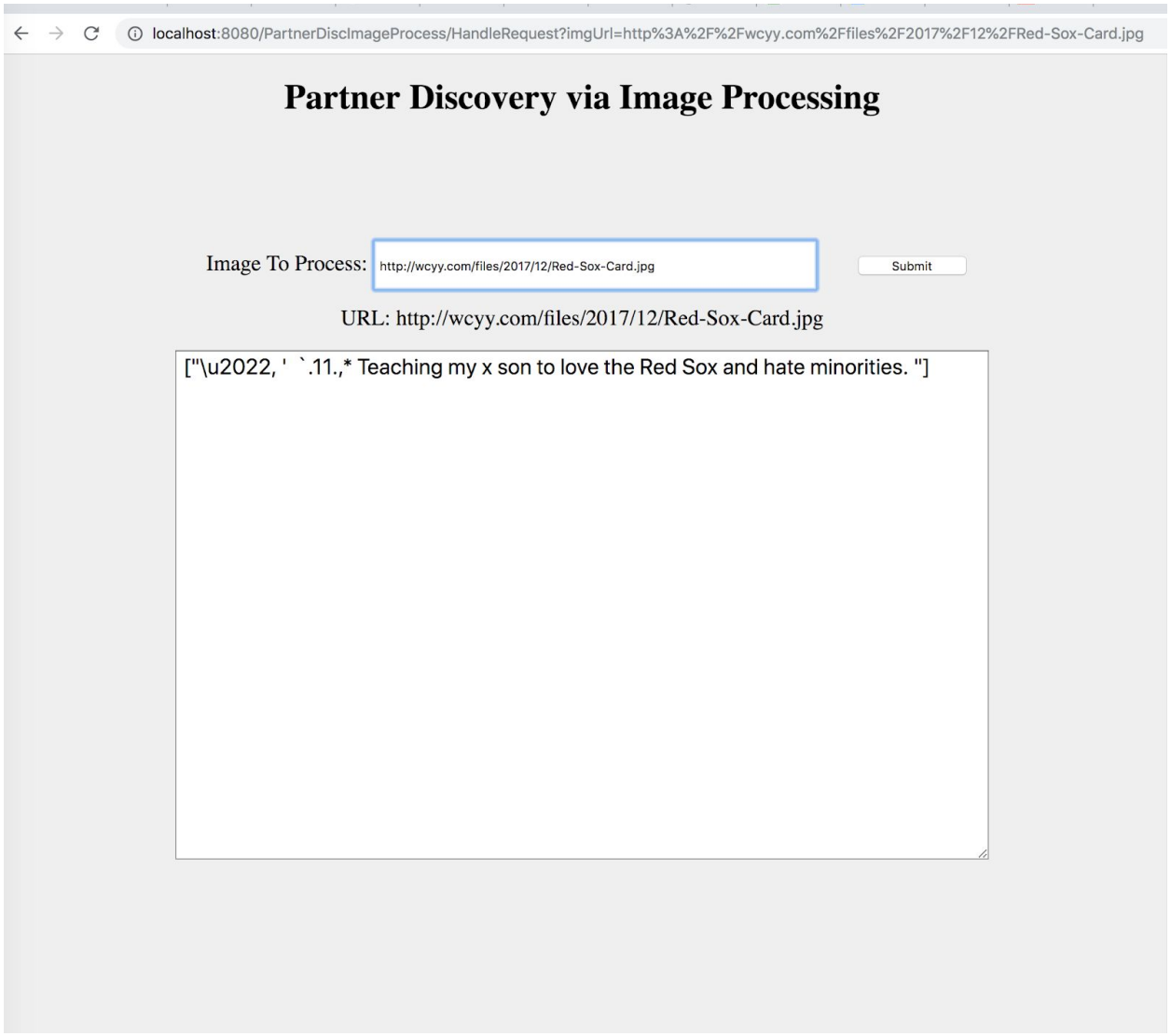
Resulting Artifacts:

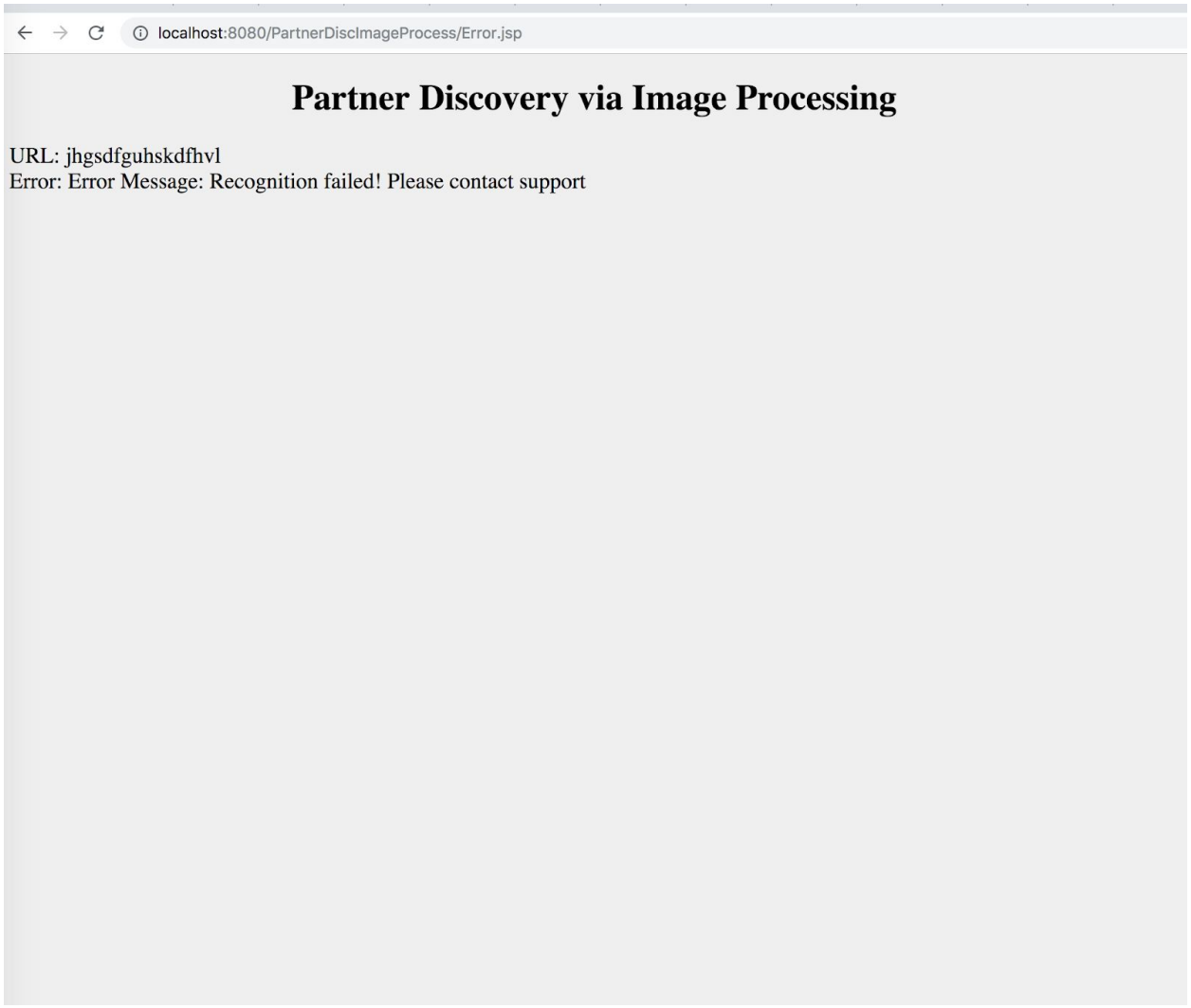
The home page is at this location
http://localhost:8080/PartnerDiscImageProcess/

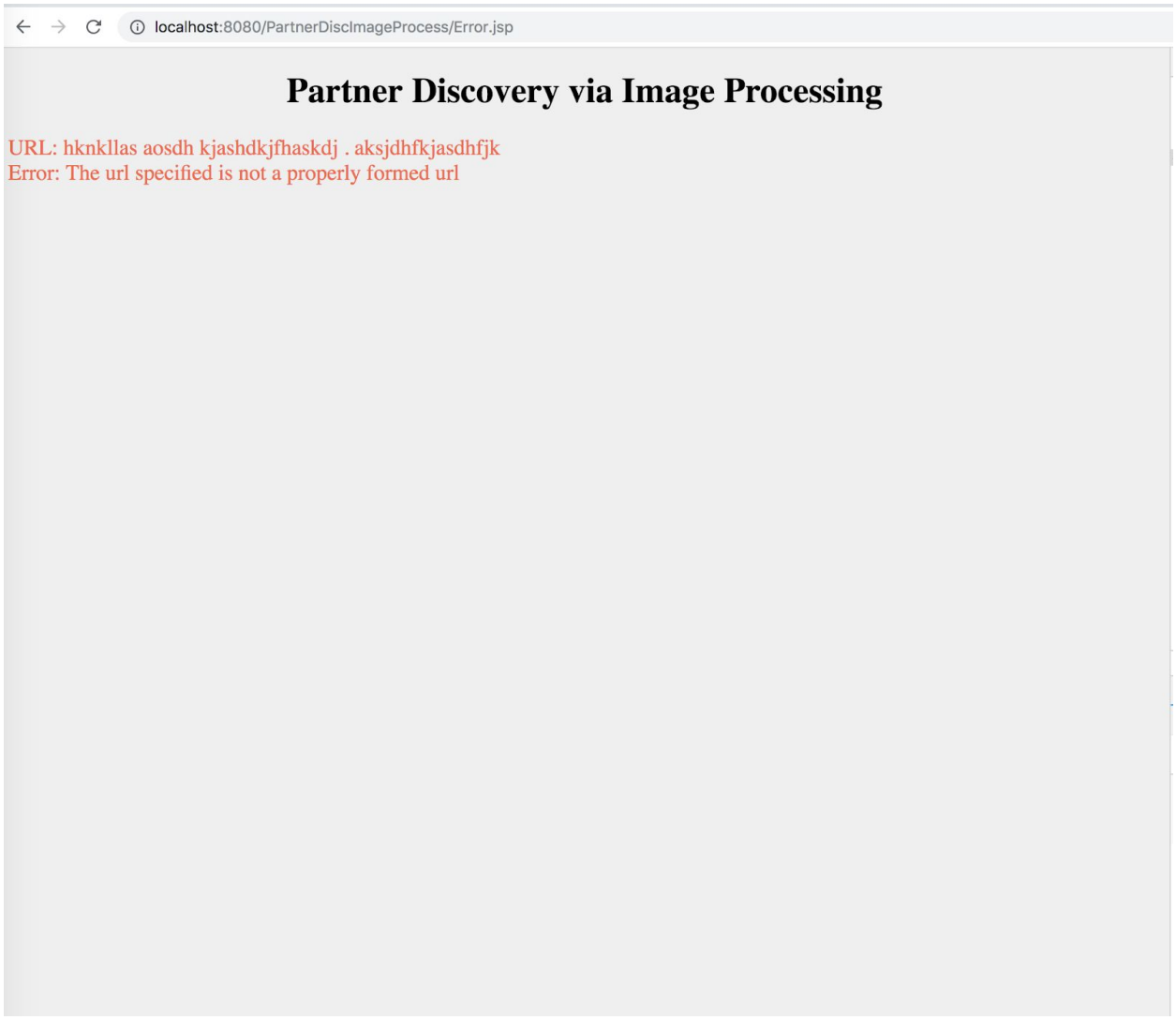
On click the index.jsp comes up with a Text box for input, a submit button and text area which displays the raw text.

There are three Java classes, HandleRequest.java, ImageProcess.java, Validator.java
There are two jsp pages, index.jsp, Error.jsp
One style page, styles.css

Below are the screenshots of the different states of the page.


localhost:8080/PartnerDiscImageProcess/HandleRequest?imgUrl=http%3A%2F%2Fwcyy.com%2Ffiles%2F2017%2F12%2FRed-Sox-Card.jpg
Partner Discovery via Image Processing
Image To Process:
http://wcyy.com/files/2017/12/Red-Sox-Card.jpg
Submit
URL: http://wcyy.com/files/2017/12/Red-Sox-Card.jpg
["\u2022, ' `.11.,* Teaching my x son to love the Red Sox and hate minorities. "]


localhost:8080/PartnerDiscImageProcess/Error.jsp
Partner Discovery via Image Processing
URL: jhgsdfguhskdfhvl
Error: Error Message: Recognition failed! Please contact support


localhost:8080/PartnerDiscImageProcess/Error.jsp
Partner Discovery via Image Processing
URL: hknkllas aosdh kjashdkjfhaskdj . aksjdhfkjasdhfjk
Error: The url specified is not a properly formed url