Plagiarism Detector
Phase B: System Design

Group 2

● ● ●

Shagun Bhardwaj
Kenji Fujita
Mason Leon
Prakash Tarun Kumar

Login UI

Plagiarism **Detector**

Username

Password

Remember Me   SIGN IN

Forgot Password?

Upload Interface

Plagiarism Detector

John Doe
john.doe@example.com

MAIN NAVIGATION

Home

Upload Files

Result

Upload Files

-- Select Language --

Upload First File.
Choose Files   No file chosen

Upload Second File.
Choose Files   No file chosen

Submit

Results Interface

Plagiarism Detector

John Doe
john.doe@example.com

MAIN NAVIGATION

- Home
- Upload Files
- Result

Results:

Function calculateFinal was found to be same at Line 45 in sample.c and demo.c

File name are different

Variable names appeared to be changed at Line 10 in sample.c and demo.c

OuputFunction is different for sample.c and dmeo.c

First Upload:

--Select File--

```
/* It will be work only IE10 */
html.ie10 .menu .list {
    background-color: #CC0000;
}
/* It will be work only Google Chrome */
html.chrome .right-sidebar .header {
    font-size: 15px;
    font-weight: bold;
}
html.chrome .right-sidebar .header {
    font-size: 15px;
    font-weight: bold;
}
html.chrome .right-sidebar .header {
    font-size: 15px;
    font-weight: bold;
}
```
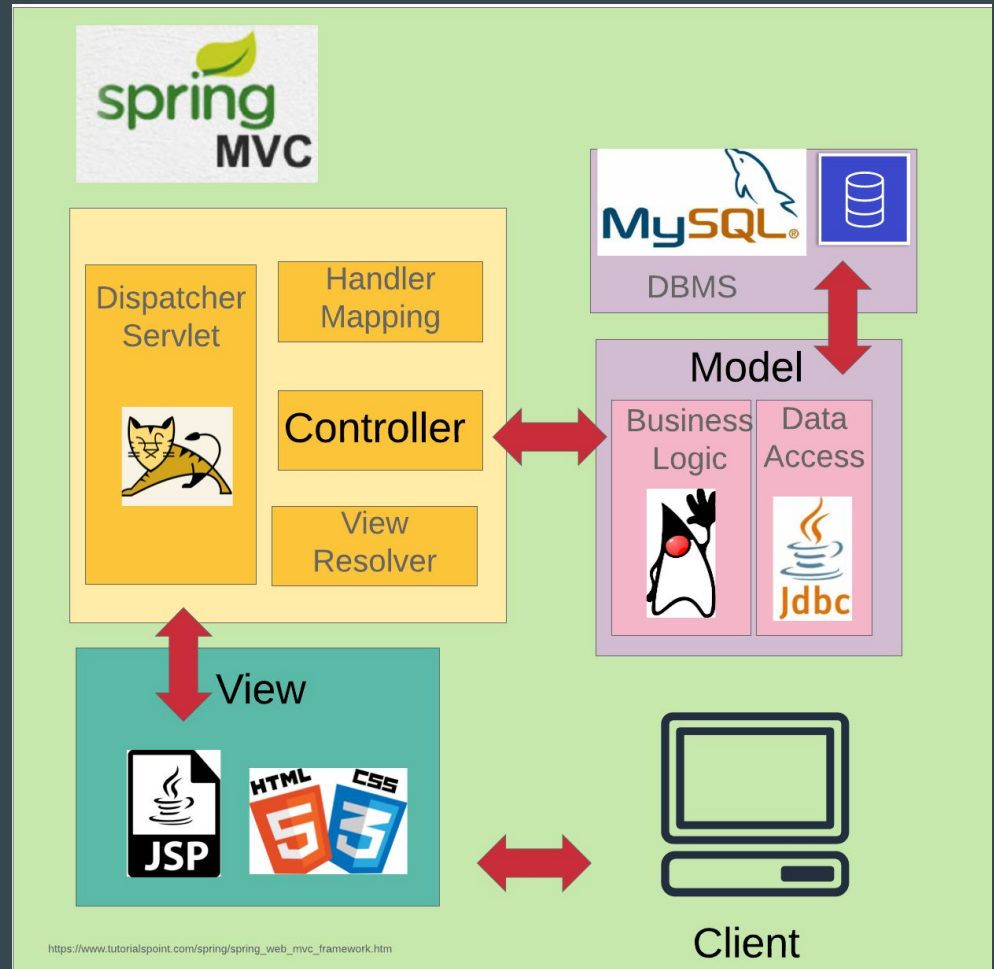
Second Upload:

--Select File--

```
/* It will be work only IE10 */
html.ie10 .menu .list {
    background-color: #CC0000;
}
/* It will be work only Google Chrome */
html.chrome .right-sidebar .header {
    font-size: 15px;
    font-weight: bold;
}
```
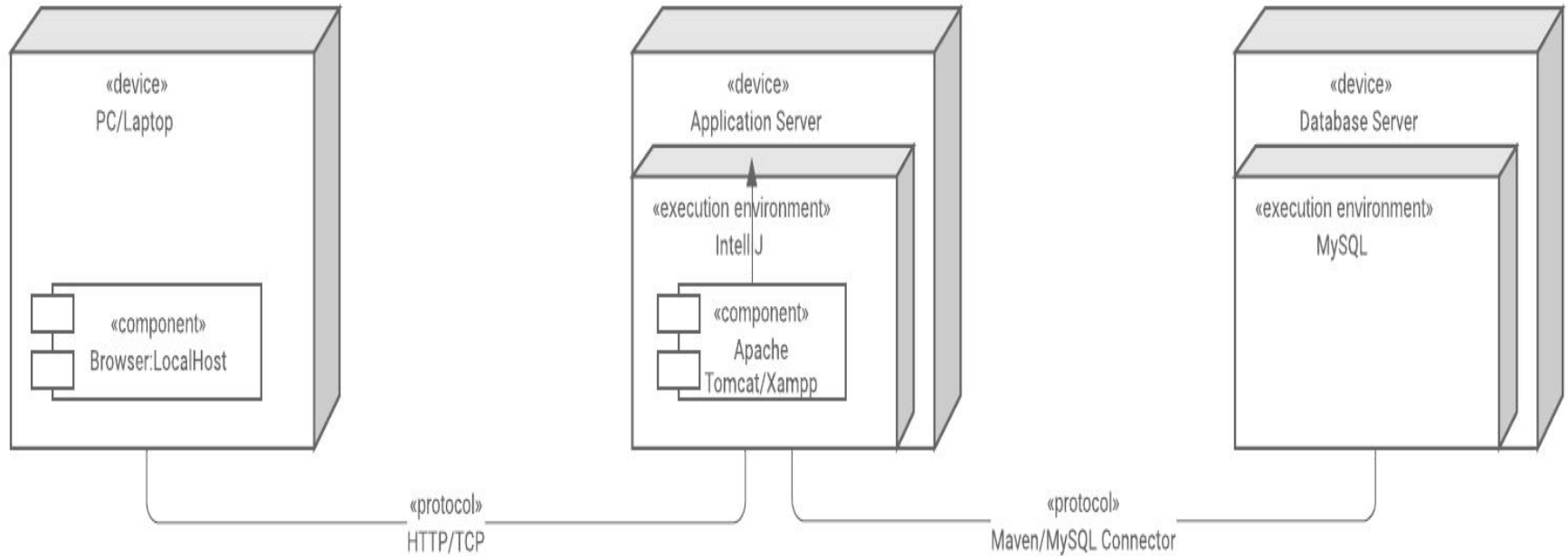
System Architecture

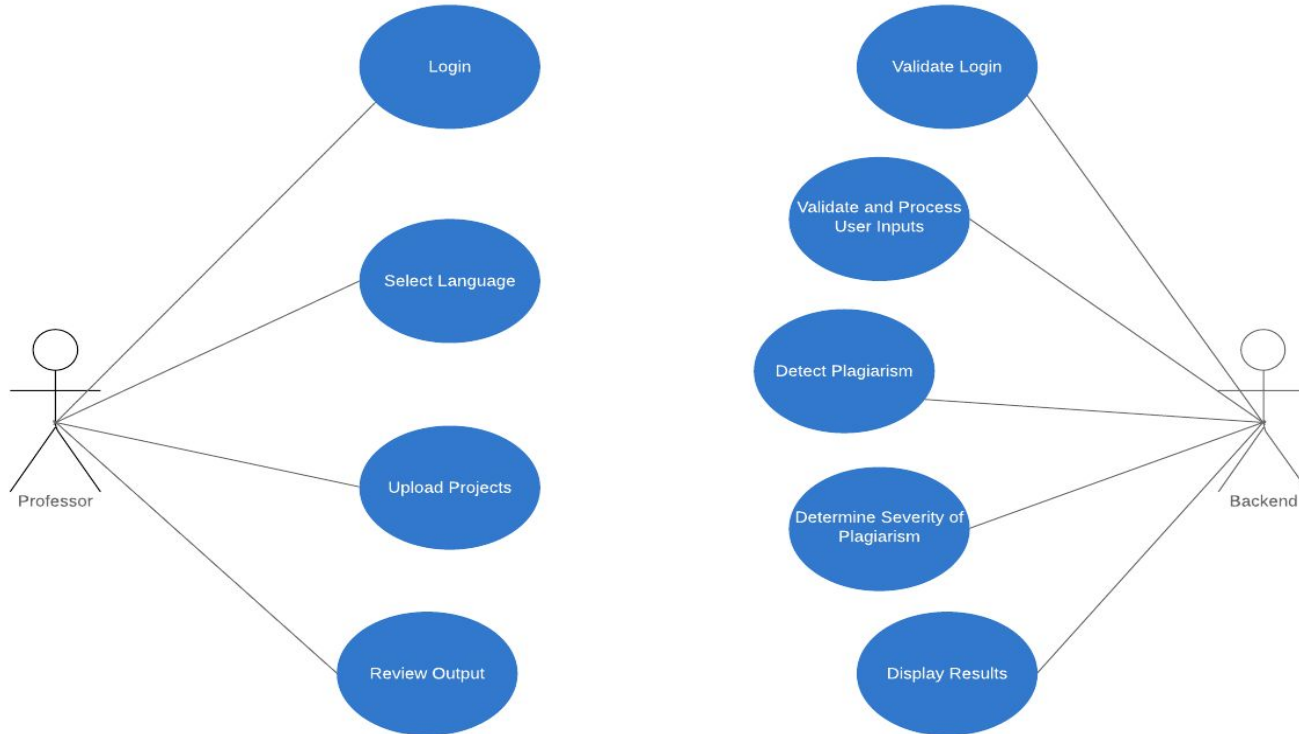https://www.tutorialspoint.com/spring/spring_web_mvc_framework.htm
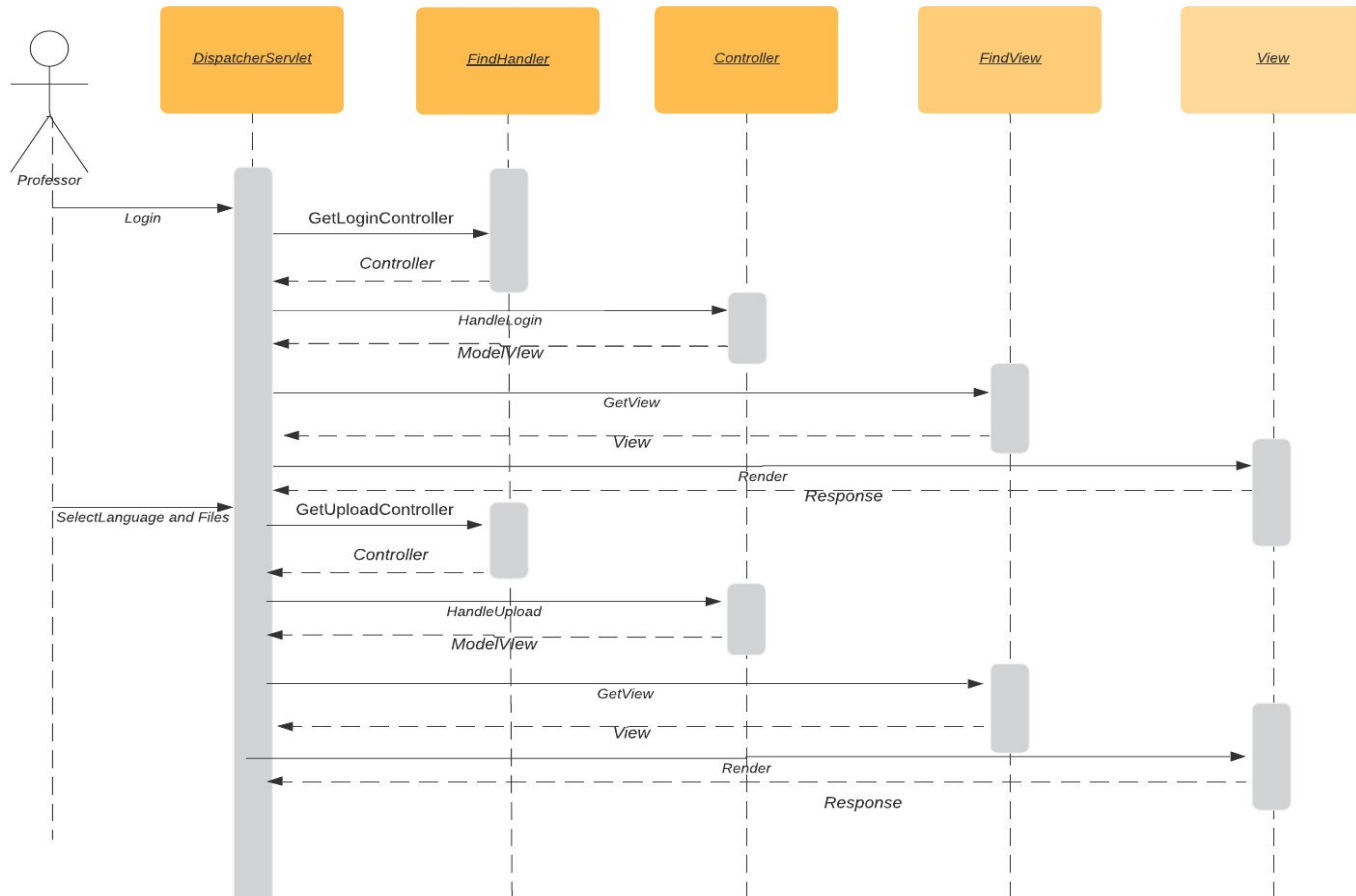
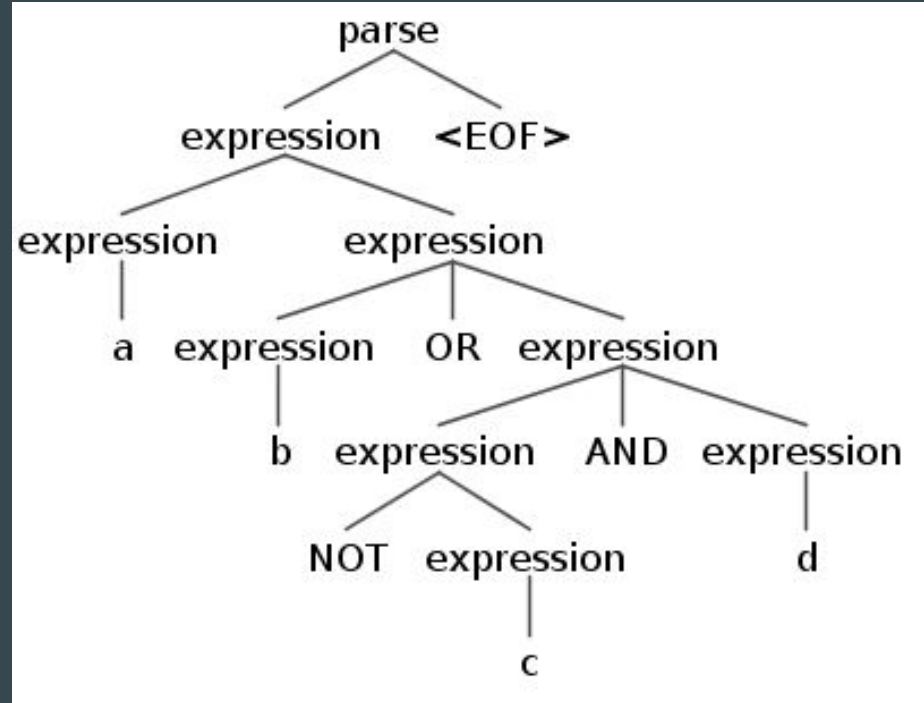# UML Deployment Diagram

UML Use Case Diagram

UML Sequence Diagram

Data Structure

- Data is parsed using ANTLR (ANother Tool for Language Recognition)
  - open source software for parsing code
- Data is stored in ParseTree object
  - Class created by ANTLR
- AST has methods to navigate tree and get data

UML Class Diagram

Model

- ParseTree tree1
- ParseTree tree2
- File package1
- File package2
- JSON plagarismdata

+ run(): void
+ analyze(): void
+ getTextComparison(): JSON
+ getGraphicComparison(): JSON
+ getMetadata(): JSON
+ getOriginal(ENUM pacakgeNum): File
+ getRawModelDS(ENUM pacakgeNum): ParseTree

Algorithms Considered

- String pattern matching
  - Compares words/characters, not semantics
  - Finds longest common subsequence
- Tokenisation
  - Pre-processes source code before comparing strings
  - Stores symbols in tokens
- Document Fingerprinting
  - Gives each document unique fingerprint
  - Similar documents will have similar fingerprints
  - Uses *k-grams* (substring length k)