

## AKNOWLEDGEMENT

First of all, our grateful goes to Allah SWT, the Most Gracious and the Most Merciful. We are indebted to our family for their patience, prayers and dua. We hope this work and degree would pay all of your sacrifices and understanding.

We also extend my deepest gratitude to our supervisor, **Dr Ayyaz Hussain** for his sustained guidance and supervision. He is always ready to respond our enquiries and spare his valuable time to correct and comment our works until the end of my study. His suggestions and improvements have guided us to come up with the best analyses and output of the dissertation.

Last but not least, we would like to thanks our friends that are always with us in tough times.

May Allah reward all of us with His mercy and forgiveness and “barokah” in this Life and the Hereafter, Amin

For everything that is right, credit goes to all of the above. For anything that is wrong, we are responsible.

## PROJECT IN BRIEF

|                        |   |
|------------------------|---|
| <b>Project Title:</b>  | Web based Document Clustering and Retrieval   |
| <b>Objective:</b>      | The main purpose of the web based app for document clustering and retrieval is to implement the document cluster techniques to categorize the documents and documents retrieval |
| <b>Undertaken By:</b>  | Majid Liaquat<br>1812/FBAS/BSSE/F12-C<br>Faran Muhammad<br>1858/FBAS/BSSE/F12-C   |
| <b>Supervised By:</b>  | Dr Ayyaz Hussain  |
| <b>Date Started:</b>   | 01-March-2016   |
| <b>Date Completed:</b> | 20-August-2016  |
| <b>Tools used:</b>     | Eclipse<br><br>Data base (My Sql)<br><br>Server (Xampp)<br><br>Operating System Windows 7 HP  |

## **ABSTRACT:**

The Purpose of this document is to collect, analyzes and defines high level needs and features of web based document clustering and retrieval. In this Document we discuss about the clustering and retrieval of the document on the basic of similarity measurement. User can share their documents and other users can get it. The cluster of documents is also created by program automatically which will depend upon the documents attributes/property. Each document is stored in different cluster after the approval of administration. For users who are willing to search the documents of their interest can be able to search the document the cluster technique will retrieve all the related documents from server regarding to user query. Query will base on two types of categories query by example and query without example.

## **DECLARATION:**

We hereby declare that this System, neither as a whole nor as a part there of has been copied out from any source, it is further declared that we have developed this System entirely on the basis of our personal efforts made under the sincere guidance of my supervisor

No portion of the work presented in this report has been submitted in support of any System for any other degree or qualification of this or any other institute of learning.

Majid Liaquat

1812/FBAS/BSSE/F12-C

Faran Muhammad

1858/FBAS/BSSE/F12-C

# Table of Contents

|  |          |
|--|----------|
| <b>1. Introduction.....</b>                                  | <b>2</b> |
| <b>1.1 Web based document clustering and retrieval .....</b> | <b>2</b> |
| <b>1.2 Existing System.....</b>                              | <b>2</b> |
| <b>1.3 Problem statement.....</b>                            | <b>2</b> |
| <b>1.4 Proposed solution .....</b>                           | <b>2</b> |
| <b>1.5 Purpose .....</b>                                     | <b>2</b> |
| <b>1.6 Scope.....</b>  | <b>3</b> |
| <b>1.7 Software Technologies .....</b>                       | <b>3</b> |
| 1.7.1 Eclipse IDE .....                                      | 3        |
| 1.7.2 Jsp servlet.....                                       | 3        |
| 1.7.3 Ajax.....  | 3        |
| 1.7.4 Java script.....                                       | 3        |
| 1.7.5 Css frame work (Bootstrap) .....                       | 4        |
| <b>2. System Analysis.....</b>                               | <b>6</b> |
| <b>2.1 Functional Requirements .....</b>                     | <b>6</b> |
| <b>2.2 Non-Functional Requirement .....</b>                  | <b>6</b> |
| <b>2.3 Design Constraints .....</b>                          | <b>6</b> |
| <b>2.4 User Characteristics.....</b>                         | <b>7</b> |
| <b>2.5 User Environment .....</b>                            | <b>7</b> |
| <b>2.6 Actor Goal List .....</b>                             | <b>7</b> |
| <b>2.7 Use Case Diagram .....</b>                            | <b>8</b> |
| <b>2.8 Use Case Description.....</b>                         | <b>9</b> |

|   |           |
|---|-----------|
| 2.8.1 Upload document .....                       | 9         |
| 2.8.2 Search documents .....                      | 9         |
| 2.8.3 Approve / disapprove documents .....        | 9         |
| 2.8.4 Apply clustering techniques .....           | 9         |
| <b>2.9. Use case Description (Detailed) .....</b> | <b>9</b>  |
| 2.9.1 Upload document .....                       | 9         |
| 2.9.2 Search Document .....                       | 10        |
| 2.9.3 Approve/Disapprove Documents .....          | 11        |
| 2.9.4 Applying Cluster technique .....            | 12        |
| <b>3. Diagrams .....</b>                          | <b>14</b> |
| 3.1 Document Uploaded Successfully .....          | 14        |
| 3.2 Document upload unsuccessfully .....          | 14        |
| 3.3 Approval of Documents .....                   | 15        |
| 3.4 Disapproval of Documents .....                | 15        |
| 3.5 Applying Cluster technique .....              | 16        |
| 3.6 Search Document by example .....              | 17        |
| 3.7 Search Document without example .....         | 18        |
| 3.8 Class Diagram .....                           | 19        |
| <b>4. System Design .....</b>                     | <b>21</b> |
| 4.1 Activity Diagram .....                        | 21        |
| <b>5. Implementation .....</b>                    | <b>23</b> |
| 5.1 Tools and Technologies .....                  | 23        |
| <b>6. System Testing .....</b>                    | <b>25</b> |
| 6.1 Why System Testing is required .....          | 25        |
| 6.1.1 Black Box Testing .....                     | 25        |
| 6.1.2 White Box Testing .....                     | 25        |

|  |           |
|--|-----------|
| <b>6.2 Test Cases.....</b>   | <b>25</b> |
| 6.2.1 Verify, that user upload a valid format of a document.....                     | 25        |
| 6.2.2 Verify, that admin must be login before applying clustering on documents ..... | 26        |
| 6.2.3 Verify, that document is store in his relevant cluster .....                   | 27        |
| 6.2.4 Verify, that user search the documents by giving query .....                   | 28        |
| <b>7. Conclusion .....</b>   | <b>31</b> |
| <b>7.1 Good Features of the System.....</b>  | <b>31</b> |
| <b>7.2 Limitations of the system.....</b>  | <b>31</b> |
| <b>7.3 Future Enhancement .....</b>  | <b>31</b> |
| <b>8. User Manual .....</b>  | <b>33</b> |
| <b>8.1 Main Page (Search document by query &amp; upload for clustering).....</b>     | <b>33</b> |
| <b>8.2 Upload Document.....</b>  | <b>33</b> |
| <b>8.3 Admin Login .....</b>   | <b>34</b> |
| <b>8.4 Admin Notification.....</b>   | <b>34</b> |
| <b>8.5 Document Approval .....</b>   | <b>35</b> |
| <b>8.6 Applying Text Clustering .....</b>  | <b>35</b> |
| <b>References.....</b>   | <b>36</b> |