A Smart Framework With Agents for eLearning System

Submitted by:
Muhammad Munwar Iqbal
2012-PhD-CS-02

Supervised by:
Dr. Yasir Saleem
Associate Professor

Department of Computer Science and Engineering
University of Engineering and Technology, Lahore
A Smart Framework With Agents for eLearning System

Department of Computer Science and Engineering
University of Engineering and Technology, Lahore
Declaration

I declare that the work contained in this thesis is my own, except where explicitly stated otherwise. In addition this work has not been submitted to obtain another degree or professional qualification.

Signed: ____________________
Date: ____________________
Acknowledgments

First of all, I would like to say special thanks to Allah Almighty (Subhaana hoo Wa Ta’ ala ) who granted me all blessings and helped me at every moment of my life and enabled me to complete and present this research work. I would also like to present the Durood and Salam on the Holy Prophet Hazrat Muhammad (Salallahoo iliahe Waa Aa’alehi Wassalam ), who taught us to walk on the right way of Islam and taught us the precious rules to live the beautiful life.

Moreover, I would also like to thank and appreciate my dear friends who supported and gave me the environment of happiness and enjoyment throughout the study period. I would also like to thanks, hostel fellows Resident tutors, Attendants, Room Servants
THIS THESIS IS DEDICATED TO ALLAH ALMIGHTY

(Subhaana hoo Wa Ta’ ala ),

HAZRAT MUHAMMAD (Salallahoo iliahe Waa Aa’alehi Wassalam ),
Contents

Acknowledgments iii
List of Figures ix
List of Tables x
Abbreviations xi
Abstract xii

1 INTRODUCTION 1

1.1 Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
  1.1.1 Electronic Learning . . . . . . . . . . . . . . . . . . . . . . . . . . 3
  1.1.2 Learning Management System (LMS) . . . . . . . . . . . . . . . . 3
  1.1.3 Moderated Discussion Boards (MDBs) . . . . . . . . . . . . . . . . 4
  1.1.4 Information Retrieval . . . . . . . . . . . . . . . . . . . . . . . . . 4
  1.1.5 Information Extraction . . . . . . . . . . . . . . . . . . . . . . . . 5
  1.1.6 Intelligent Agents . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
  1.1.7 Question Answering . . . . . . . . . . . . . . . . . . . . . . . . . . 5
1.2 Motivations towards Study . . . . . . . . . . . . . . . . . . . . . . . . . 6
1.3 Research Gap . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
1.4 Approach of the Study . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
1.5 Problem Statement . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
  1.5.1 Research Questions . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
1.6 Aims and Objectives of the Study . . . . . . . . . . . . . . . . . . . . . . 11
1.7 Significance of Proposed Study . . . . . . . . . . . . . . . . . . . . . . . . 12
1.8 Scio-Economics Benefits . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
1.9 Research Contributions . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
1.10 Thesis Road Map and Conventions . . . . . . . . . . . . . . . . . . . . . 15

2 LITERATURE REVIEW 16

2.1 Background . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16
  2.1.1 Factors Increasing eLearning Enrollments . . . . . . . . . . . . . . 17
2.2 Course Offered for the eLearning Paradigm . . . . . . . . . . . . . . . . 17
  2.2.1 Top 10 eLearning Websites . . . . . . . . . . . . . . . . . . . . . . . 18
  2.2.2 Learning Resources . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
  2.2.3 Online Learning . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
2.3 Web Semantics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
Contents

2.3.1 The Current Web ........................................... 20
2.3.2 Web 2.0 .................................................. 21
2.3.3 The Semantic Web ....................................... 21
2.3.4 Modeling Ontology ...................................... 21
2.3.5 Ontology Development .................................. 23
2.3.6 Semantic Web Services ................................. 23
2.4 Agent Technology ........................................... 25
2.5 Agents Types ................................................ 26
  2.5.1 Mobile Agents .......................................... 26
  2.5.2 Collaborative Agents .................................. 26
  2.5.3 Reactive Agents ........................................ 27
  2.5.4 Interface Agents ...................................... 27
  2.5.5 Hybrid agents ......................................... 27
  2.5.6 Internet / Information Agents ....................... 27
  2.5.7 Service Agents ....................................... 27
  2.5.8 Intelligent Agents .................................... 27
2.6 Multi-Agent System ........................................ 28
2.7 Agent Application Domains ................................ 29
2.8 Agent Development Environments ....................... 29
  2.8.1 JAVA Agent Development Environment (JADE) .... 29
  2.8.2 FIPA-OS ............................................... 30
  2.8.3 Agent Building Ad Learning Environment(ABLE) .. 30
  2.8.4 AgentBuilder ......................................... 30
  2.8.5 Aglets ................................................ 30
  2.8.6 Java Agent Template Lite (JATLite) ................ 30
2.9 Question Answering ......................................... 31
  2.9.1 START System ......................................... 32
  2.9.2 Answer Bus ............................................ 33
  2.9.3 Language Computer Corporation (LCC) ............ 33
  2.9.4 Answefinder ......................................... 33
  2.9.5 Question Answer System for the Bengali Language 34
  2.9.6 Chinese Question Answer System .................... 34
2.10 Summary of Literature Review ............................. 35

3 DELAY ASSESSMENT IN QUESTION ANSWERING SYSTEM 38
3.1 Question Answering in LMS ................................. 38
3.2 Delay Factor in Existing System ......................... 39
  3.2.1 Data Acquisition .................................... 40
  3.2.2 Data Pre-Processing ................................ 40
  3.2.3 Data Analysis ....................................... 41
3.3 Algorithm Pseudo Code for the Delay Data Analysis .... 42
  3.3.1 Trends in Wordcloud ................................ 44
3.4 Statistical Analysis of the Collected Data ................ 45
3.5 Students Cluster versus Delay in Hours ................. 46
3.6 Number of Records versus Sum of Delays ................ 47
3.7 Delay Against Distinct Student IDs ....................... 48
3.8 Correlation Representation .............................. 48
3.9 Conclusion ................................................................. 49

4 MULTIMEDIA AND AGENT BASED QUESTION ANSWERING SYSTEM FOR ELEARNING 50
4.1 Question Answering Architecture ............................ 50
4.2 Proposed “Question Answering (QA) System” Architecture 51
4.3 Development Tools .................................................. 55
4.3.1 Python ............................................................. 56
4.3.2 WordNet .......................................................... 56
4.3.3 Natural Language Toolkit (NLTK) ......................... 56
4.3.4 Alchemy Keyword Extraction API ......................... 56
4.4 Multi-agent Object Oriented Question Answering (MOQA) 56
4.5 Student Question Posting Interface ......................... 61
4.6 Student Question Categorization ............................ 61
4.6.1 Text Questions .................................................. 63
4.6.2 Image Questions ................................................ 63
4.7 Question Correction Module (QCo Agent) ............... 64
4.7.1 Spelling Correction Types .................................. 66
4.7.1.1 NonWord Correction ..................................... 66
4.7.1.2 RealWord Correction ..................................... 66
4.7.2 NOISY CHANNEL Spelling Correction ................. 67
4.7.3 Keywords Extraction Module ............................. 70
4.7.4 Sentence Tokenization Process ........................... 72
4.7.5 Removal of Punctuation Marks ......................... 72
4.7.6 Part of Speech (POS) Tagging ............................ 72
4.7.7 Parsing ............................................................. 73
4.7.8 Named Entities (NE) ........................................... 73
4.8 Keywords Extraction Module (KE Agent) ............... 74
4.8.1 Keywords Extraction .......................................... 74
4.8.2 Headword Extraction ........................................ 74
4.9 Question and Answer Type Identification ............... 75
4.9.1 Question Classification Algorithms .................... 75
4.9.2 Question Answer Types ...................................... 76
4.9.3 Definition Type Questions ................................. 82
4.9.4 Description Type Questions ............................... 82
4.9.5 Factoid Type Questions ..................................... 82
4.9.6 Image based Type Questions ............................. 84
4.10 Answer Extraction and Processing Module ............. 84
4.10.1 Web Documents Retrieval ................................. 84
4.10.2 Google ........................................................... 85
4.10.3 Wikipedia ......................................................... 85
4.10.4 Webopedia ....................................................... 86
4.10.5 Answers Extraction Module ............................. 87
4.10.6 Different Corpus Text Files ............................. 88
4.10.7 Extracted Text Data Documents ...................... 89
4.10.8 Remove Hyperlinks and Hash Tags ................... 90
4.10.9 Documents Processing Module ......................... 90
4.10.10 Sentence Tokenization Process ...................................... 90
4.10.11 Training Process for Part of Speech (POS) Tagging .............. 90
4.10.12 Part of Speech (POS) Tagging ...................................... 92
4.10.13 Parsing Sentences to each Extracted Text Document .............. 92
4.10.14 Finding Named Entities (NE) ....................................... 93
4.10.15 Named Entities Categories .......................................... 93
4.11 Extracting Keywords and Relevancy Score Module ....................... 94
4.11.1 Extracting Keywords .................................................. 95
4.11.2 Finding Relevancy Score of Keywords Techniques ............... 95
4.11.3 Answers Extraction ................................................... 95
4.11.4 Answers Extraction Algorithms ...................................... 98
4.11.5 Extracted Answers from Extracted Text Documents ............... 102
4.11.6 Answers Ranking Techniques ........................................ 102
4.11.7 Candidate Answers ................................................... 106
4.12 Image based Answers Module (IA Agent) ................................ 106
4.12.1 Image question Processing .......................................... 106
4.12.2 Student Ranking Answer ............................................. 109
4.12.3 GPS Coordinates Collection ........................................ 111
4.12.4 Google QA System ................................................... 111
4.12.5 User Questions and Extracted Answers Repository ............. 111
4.13 Case Study ..................................................................... 114
4.13.1 Keywords Relevancy in Answers Response to Student Question 114
4.13.2 Students Collected Data through Smartphone GPS .................. 115
4.14 Appointment Fixing Module ............................................. 117
4.14.1 Multimedia Answer Contents to Student ................................ 118
4.15 Conclusion .................................................................... 119

5 RESULTS AND DISCUSSIONS .................................................. 120
5.1 Performance Evaluation ..................................................... 120
5.1.1 Parameters for evaluation ............................................. 121
5.1.2 Recall Measurement .................................................... 121
5.1.3 Precision Measurement ................................................ 122
5.1.4 Accuracy Score Measurement ........................................ 122
5.1.5 F-Measurement .......................................................... 122
5.1.6 Specificity Measurement .............................................. 123
5.1.7 Matthews Correlation Coefficient .................................... 123
5.2 Multimedia and Agent based Question Answering System Analysis .......................................................... 123
5.3 Accuracy Comparison for the Image Type Questions .................... 124
5.4 Comparison of the QA System in terms of Response and Accuracy .... 125
5.5 Analysis of MAQAS and Teacher Answers ................................ 126
5.6 Conclusion .................................................................... 132

6 CONCLUSION AND FUTURE WORK ...................................... 134
6.1 Conclusion .................................................................... 134
6.1.1 Next Generation Question Answering Systems .................... 134
6.1.2 Delay Assessments ...................................................... 135
6.1.3 Multimedia and Agent-Based Question Answering System .... 135
6.2 Future Work ................................................................. 136
  6.2.1 Blended Learning ................................................... 136
  6.2.2 Virtual Reality .......................................................... 136
  6.2.3 Augmented Reality .................................................... 136

7 REFERENCES ................................................................. 138
List of Figures

1.1 Research Gap Identification for this Dissertation .......................... 7
1.2 Student and Instructor Collaboration in a Pure eLearning Setup ........... 9
2.1 Course Distribution by Discipline for the eLearning Paradigm ............. 18
2.2 Course Length and Monthly Availability .................................. 18
2.3 The Semantic Web Stack (from W3C) ....................................... 22
3.1 Student Interaction via Learning Management System ........................ 39
3.2 Student QuestionAnswering MDBs Interface ................................ 40
3.3 Data Analysis Model for Delay Measurement in the QA System .......... 41
3.4 Word Cloud Trends in Question Text ....................................... 45
3.5 Students Cluster versus Delay in Hours ................................... 46
3.6 Number of Records versus Sum of Delays .................................. 47
3.7 Delay Against Distinct Student IDs ....................................... 48
3.8 Student’s Question Terms Correlation with Overall Dataset ................ 49
4.1 Block Diagram of Proposed Model for eLearning Setup .................... 52
4.2 The Proposed Model for eLearning Setup .................................. 53
4.3 Multimedia and Agent based Question Answering System (MAQAS) ....... 54
4.4 Architecture of Multi-agent Object Oriented Question Answering (MOQA) 57
4.5 Gaia Design and Analysis Model for the Agents [2094] ...................... 58
4.6 Generalized Architecture of the System ................................... 60
4.7 Question Posting Interface for Students ................................... 60
4.8 Question Categorization Module .......................................... 61
4.9 Question Categorization Interface ......................................... 63
4.10 Agent Collaboration in Question Answering System ....................... 64
4.11 Question Statement Correction Module ................................... 65
4.12 Keywords and Headwords Extraction ...................................... 71
4.13 Sentence Tokenizing and Parsing ....................................... 72
4.14 Question Part of Speech Annotation ...................................... 73
4.15 Sentence Parsing ............................................................. 73
4.16 Question Type Classification ............................................... 76
4.17 Key Words Extraction Process from Text Data ............................. 77
4.18 Question Classification using Question Words ............................. 86
4.19 Answer Type Identification using Question Words Classification ........ 87
4.20 Top 10 URls from Google Search Engine .................................. 88
4.21 Data Extracted from Google Search Engine ................................. 91
4.22 Data Extracted from Wikipedia Search Engine .............................. 91
4.23 Answer Sentence from the Text Extracted and Tokenized .......................... 92
4.24 Part of Speech Tagging for Answering Sentence ................................. 92
4.25 Named Entities Identified by Entity Extractor Tool ............................... 93
4.26 Extracted Sentence for Answer .......................................................... 94
4.27 Named Entities (NE) Identified from Answer ........................................ 94
4.28 Answer Extraction and Processing Module ........................................... 96
4.29 Key Words Extraction Process from Text Data ..................................... 97
4.30 Image base Question Processing for Answers ....................................... 107
4.31 Image Annotator Tool for the Annotation and Validation of Images .......... 108
4.32 Answer Delivery and Appointment Module .......................................... 110
4.33 Answer Extracted from Google QA Snippets ....................................... 112
4.34 Repository for MAQAS System ............................................................ 113
4.35 Answer for Repository of MAQAS System ........................................... 114
4.36 A Subset of the Data of Student Cluster on Google Map .......................... 117
4.37 Multimedia Content Delivery Model .................................................... 118

5.1 Comparison of the QA System in terms of Response and Accuracy ............. 124
5.2 Accuracy Comparison of MAQAS with Online Systems .......................... 126
5.3 Comparison of the MAQAS with Photo Based QA systems ....................... 127
5.4 Features based Comparsion of MAQAS and Existing QA Systems ............. 127
5.5 Comparison of MAQAS with Teacher Answers ....................................... 129
## List of Tables

2.1 Factors Increasing eLearning Enrollments ........................................... 17  
2.2 Top 10 eLearning Websites and Resources for the Students .................. 19  
2.3 New trends in the Web Concept ........................................................... 21  
2.4 Comparison of Different Techniques among Question Answering Systems .... 35  
2.5 Summary of Literature Review ............................................................. 37  
3.1 Descriptive Statistics of the Dataset ..................................................... 45  
4.1 Responsibilities and Activities of LMS Agent ......................................... 58  
4.2 Responsibilities and Activities of QA Agent ........................................... 62  
4.3 Responsibilities and Activities of QC Agent ........................................... 62  
4.4 Responsibilities and Activities QuestionCorrection Agent ......................... 65  
4.5 Candidate Corrections for the Misspelled Word ”huble” .......................... 67  
4.6 Candidate Corrections for the Misspelled word ”futher” ........................... 68  
4.7 Word with Reasonable Weight Estimation ............................................. 68  
4.8 Probabilities for the Words Insertion, Deletions, Transposition and Substitutions ................................................................................................................. 69  
4.9 Candidate Corrections for the Misspelling Across and the Transformations .......... 69  
4.10 Ranking of Each Candidate Correction using Language Model ................. 70  
4.11 Responsibilities and Activities of KeywordExtraction Agent ..................... 72  
4.12 Responsibilities and Activities of QuestionAnsweringIdentification Agent .... 75  
4.13 Categorization of WH-Word Type Question ........................................... 76  
4.14 Classification and Question Focus of Different Question Words or Phrases .... 83  
4.15 Responsibilities and Activities of AnsweringExtraction&Processing Agent .... 85  
4.16 The Determiner Terms Corpus Text File ................................................ 88  
4.17 The Definitions Terms Corpus Text File ................................................ 89  
4.18 The Reasoning Terms Corpus Text File ............................................... 89  
4.19 Responsibilities and activities of FeedbackRelevancy Agent ...................... 102  
4.20 Keywords Relevancy Calculations in Answers to Student ......................... 115  
4.21 A Subset of Students Collected Data through Smartphone GPS ................ 116  
4.22 Responsibilities and Activities of AppointmentFixing Agent .................... 116  
5.1 Performance Analysis with other Question Answering Systems ................. 124  
5.2 MAQAS vs Photo-Based QA System Image Questions ............................. 125  
5.3 MAQAS Comparison with Teacher Answers ........................................... 128  
5.4 Comparison of Techniques MAQAS with QA Systems ............................ 130  
5.5 Comparison of MAQAS Features with Existing QA Systems .................... 132
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOQA</td>
<td>Multi-agent Object-Oriented Question Answering</td>
</tr>
<tr>
<td>MAQAS</td>
<td>Multimedia Agent based Question Answering System</td>
</tr>
<tr>
<td>NLP</td>
<td>Natural Language Processing</td>
</tr>
<tr>
<td>NLTK</td>
<td>Natural Language Tool Kit</td>
</tr>
<tr>
<td>LMS</td>
<td>Learning Management System</td>
</tr>
<tr>
<td>MDBs</td>
<td>Moderated Discussion Board</td>
</tr>
<tr>
<td>QA</td>
<td>Question Answering</td>
</tr>
<tr>
<td>NE</td>
<td>Named Entity</td>
</tr>
<tr>
<td>NER</td>
<td>Named Entity Recognizer</td>
</tr>
<tr>
<td>IDLE</td>
<td>Integrated Development and Learning Environment</td>
</tr>
<tr>
<td>IR</td>
<td>Information Retrieval</td>
</tr>
<tr>
<td>POS</td>
<td>Part of Speech</td>
</tr>
<tr>
<td>START</td>
<td>SynTactic Analysis utilizing Reversible Transformations</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>IE</td>
<td>Information Extraction</td>
</tr>
<tr>
<td>DM</td>
<td>Data Mining</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>API</td>
<td>Application Programmable Interface</td>
</tr>
<tr>
<td>TF</td>
<td>Term Frequency</td>
</tr>
<tr>
<td>IDF</td>
<td>Inverse Document Frequency</td>
</tr>
<tr>
<td>NER</td>
<td>Named Entity Recognizer</td>
</tr>
</tbody>
</table>
Abstract

Electronic Learning (eLearning) has attained significance popularity all over the world due to its diverse nature. This trend introduced the need for intelligent systems based on agents in order to respond to some certain levels of student’s questions. Presently, a large number of students enroll eLearning courses, and instructors cannot accommodate the high volume of student’s questions. The learning experience is severely affected by delays in responding to student’s questions. To address this delay, computer-aided question-answering systems are needed. Web technology and multi-agent tools are combined to provide a service-oriented system. The core purpose is to deliver the augmented solution to the submitted student’s questions. The Multi-Agent Object-oriented Question Answering (MOQA) architecture is designed to work with a Learning Management System (LMS). MOQA is the multi-agent system that used to communicate between Question Answering and LMS to facilitate the student. Multimedia based Agents-oriented Question Answering System (MAQAS) handled the questions posted by a student. The posted questions are first parsed for processing, in which a complete, concrete understanding of the question is achieved using intelligent agents. This system searches and discovers all possible semantic paths that can provide a possible solution. By keeping the student’s question in view, agents use specific techniques to find a precise answer. General questions are answered in a short time, similar to a live answer. Other questions are narrowed down to a more specific level using few intelligent questions. The relevancies of the answers are calculated along with feedback from student to check the student satisfaction. This research enhances the online learning capabilities for more students concurrently and manages the knowledge base for future usage.
Chapter 1

INTRODUCTION

This chapter introduces the fundamental concepts of the work presented in this dissertation. Electronic Learning (eLearning) technology and its contribution in the field of Science and Social Sciences education will be discussed. The objective of this research and the problems faced by the students in the eLearning environment are also discussed. The major tools used in carrying out this research are the; agent with NLTK, WordNet API and Python as development language with Java. The organization of the thesis is also presented at the end of this chapter.

1.1 Overview

The Electronic Learnings integration in the education system is embraced and acknowledged globally. eLearning enhance the student learning curve and satisfaction by using processes and learning technologies. These learning technologies are used to acquire mandatory skills and expertise in the educational field. eLearning is the paradigm in which knowledge acquisition and transfer are characterized by its effectiveness [1]. There are a number of reasons for the adaptation of using information and communication technologies (ICT) for the learning purposes in education and some of them may be inadequate [2]. Students are used to Learning Management System (LMS) integrated with Moderated Discussion Boards (MDB) for querying their questions and answers [3]. The motivation and constructivism in the learning environment are directly related to each other that enhance the students learning. Students self-efficacy also in the eLearning environment increases the learning curve positively [4]. While the student’s questions unanswered in the eLearning environment create anxiety in students, it slows down the
student learning, especially in an eLearning environment. eLearning is an evolving field that attained popularity in the recent decade. In eLearning paradigm, courses and programs are offered via internet based resources [5].

The evolution of eLearning is result in the development of information and communication technologies (ICT) in education. ICT is used to facilitate a large number of the student to learn and get trained [6]. It minimized the gap between the education resources and the students. eLearning differs from the traditional learning as it can be used in both types: synchronous and asynchronous modes of communication. But eventually, at the same time increase the distance between the teacher and student. A lot of stuff is available on the internet for the online learners [7]. Massive Open Online Courses (MOOCs) are also available for the learners and the students [8]. MOOCs are designed to teach the single topic content. MOOCs have very short time duration as the video lecture. The assessment of the delay is needed to be calculated for the betterment of the student [9]. It leads to minimizing the gap between the student-teacher interaction. The analysis of the question text will be a help to find the weak area while answering the student question.

Cloud computing virtualization of the resource technologies and innovation suitability for the participants are the key difficulties in adopting the recent learning advancement to respond to the participants of the eLearning. Although Learning Management System acceptance and adaptation is increasing rapidly, the problem faced by the students, teacher and administrators from different perspectives need to be addressed: administrator faces the problem to retain the students, students face problem in finding answers to their questions arose during understanding their topics and lectures [10].

There are two principal aspects to these perspectives: the first is the introduction of methods for arranging the learning material; the second concerns the measurement of intelligence and potential outcomes. These features are designed and found to be feasible for the dual effort to explore potential outcomes. However, the emergence development, improvement, and across the board utilization of new correspondence and coordination efforts are realized in Web 2.0. It is therefore, imperative that eLearning becomes more social and natural for students. The student can imagine the additional benefits that Open Web innovation offers. It makes possible to the improvement of the quality of audio files. Another example is a web-driven conferencing framework that may be utilized for synchronized online coordination and collaboration. This framework uses internet technology and executed in eLearning environments progressively and continuously. There is
no face-to-face interaction between students and other stakeholders, i.e., the academic, administrative, or registrar office staff. Coordination and communication occur through web services available on university’s web server. To achieve these objectives, everyone in the eLearning system is supported by web documents and web services. An agent may be used for these web services in order to enhance communication and to provide a effective answering system.

1.1.1 Electronic Learning

A form of education and its management delivered by means of Electronic Resources and Information Technology is called Electronic Learning (eLearning). The major elements of eLearning are hardware, software applications, databases, web applications, web documents, digital media, and human resources. eLearning in contrast to traditional classroom learning involves learning from spoken or printed materials as well as from images, animated pictures and videos with the help of technologies. Electronic Learning is designed for instructors, learning technologists, and people with a general interest in education. eLearning has become a very speedy innovative paradigm by demonstrating itself as an effective educational technique [10]. Learners who want to deepen their knowledge use eLearning courses and training from international online universities. In contrast to traditional classroom learning, eLearning is defined as knowledge collaboration using electronic media. To achieve the learning goals constructivism in a learning environment is its most important objective [11]. Electronic learning is perceived as assisting, mentoring and enhancing the traditional learning styles and methods in innovative manner. It creates more opening for the student and learners at their campuses and off campuses.

1.1.2 Learning Management System (LMS)

Learning Management System (LMS) is an essential part of eLearning paradigm, which is used to provide the virtual environment for the online learners. The teachers and instructors are interacting with students through LMS [10, 12]. The education resources are delivered to the students via Learning Management System. LMS is shown Figure 1.2 with its components. The components of the eLearning system are students, teachers, LMS, Audio-video Lectures, SMS Services, Handouts, PowerPoint slides, e-mail and other download-able materials. The teachers record their video lectures which are uploaded at LMS for the access of student enrolled in the virtual universities. These
lectures are sometimes available at some openly accessible websites like YouTube, Open Course Ware and broadcast on television channels. Open Education Resources (OER) are also available for enrolled students [5, 7]. A student can also access their other learning material like handouts, reference books, tutorial, quizzes, assignments via LM-S. Students have options for their subject related matter to communicate their teacher by Moderated Discussion Boards (MDBs) and emails. These emails and MDBs questions are replied by the instructors [12].

1.1.3 Moderated Discussion Boards (MDBs)

Student support is one of the most challenging problems arising from the availability of a huge number of eLearning courses and millions of enrollments. Elements of student support include databases (DBs), web applications, web documents, software applications, hardware tools, digital media, and human resources. Students use Learning Management System (LMS) which includes Moderated Discussion Boards (MDBs) for posting questions [13]. They have an equal privilege to post a question at MDBs, which is visible to all classmates. The present technological resources used to drive eLearning are eLectures, eBooks, netbooks, iPods, and dialog-tool. The questions posted by the students are answered with service delay. This delay of the students posted questions are directly proportional to the number of the question posted at moderated discussion boards and inversely proportional to the number of the instructor available to answer the student’s questions. The moderate discussion boards are designed separately for each subject and available for registered students only. The answers to students question are available for only the student registered in the subject.

1.1.4 Information Retrieval

The Text Retrieval Conference (TREC) in 1999 focused on the theme of Question Answering (QA). The researcher’s focus was directed towards Information Retrieval (IR) in a QA framework. In this conference shifted the focus from document retrieval to direct and indirect QA [14]. This led to proposals considering the hypothesis that the user usually wants to get the answer to his question instead of a large number of documents [15]. QA entails data management, data retrieval, Natural Language Processing (NLP), and dialect handling. The goal of IR system is to identify the components and the related text. The IR system is guided towards the client’s predefined requirements. The goal of characteristic dialect handling is to establish an environment for the dialog
between the client and the QA framework using NLP [16]. The same targets of QA are obvious in a definition type of questions-answering framework.

1.1.5 Information Extraction

In TREC Conference, 2006, the focus in question answering is shifted from information retrieval (IR) to information extraction (IE). Information extraction is the process of the extraction of the text from the unstructured and semi-structured documents.

1.1.6 Intelligent Agents

An intelligent agent is a piece of code that executes and follows a defined path. The general characteristics that define intelligent agents include: a) the first is that they operate continuously, b) the second is that they have the ability to make decisions based on input information corresponding to the current situation, c) the third characteristic is that they strive to achieve the desired output, d) the fourth is their interoperability, e) the fifth is that the agent will learn from the context and user inputs [17]. Despite the continuous growth of the web, people want to find specific information without any delay. To achieve this goal, the semantic web was developed, which provides a way of adding semantics to published data. It allows machines to understand and process data in a human-like way. Ontology development is an important technique in this approach, which is used by the agent for better results [18].

1.1.7 Question Answering

Automated QA was first introduced in the 1960s, confined to closed domain expert system [19, 20]. QA systems have experimented on closed knowledge bases [21-23] and pre-tagged corpora [24-26]. These systems focus on Text Retrieval Conference tasks. In TREC 2006 focus is shifted from IR to IE for Question answering. The MAQAS is a Smart Framework that manages the experimental setup by integrating the components such as; Corpus, Multi-Agent System, Knowledge Repository, Natural Language Processing Toolkit, Image Annotator, Web Document, UML Vocabulary, Question processing module, Answer processing and ranking module, Students, and Teachers. Different types of answers relate to QA system are like factoid, description, definition, ratings. Here, an eLearning environment that combines a MAS and Semantic web technology is proposed in order to achieve better search results and task workflows. Semantic Web Services (SWS) with Multi-Agent System (MAS) and eLearning are very active areas of research. These are integrated to work and complete the automation of the solution work
flow and a task-centric environment. We will try to build the system by combining these strengths and mechanisms, and attempt to refine the functionalities of multi-agents on semantic web technology in an eLearning environment.

1.2 Motivations towards Study

Moderated Discussion Boards (MDBs) are used by different universities at national and international level. They allow students to post questions and instructors to reply to them one by one. Like Blackboard, there is an online shared-screen mechanism that is used for students assistance and guidance. Both the instructor and the students interact online and the instructor can use the MDB like a conventional whiteboard or blackboard to help the student to grasp a concept. From the perspective of the students, the manifestation of intelligent agents is valuable. Moreover, in the humanoid computer system with intelligent agents are emerging and considerably improving student’s communication.

Many eLearning staff and administrators report a chronic gap between the instructors and the students. Research gap exists for a number of reasons, including the increasing student’s demands of eLearning, a large number of eLearning courses offered by universities. It also depends on budget, which can exacerbate the gap as it can result in a reduction of number of online class sessions, and can restrict the hiring of the faculty required to support the students. Administrators in eLearning are facing many challenges in trying to provide the necessary facilities including virtual student support, online library access, materials, advisory services, and career counseling, all of which are available on campus for face-to-face university offerings. However, while a student may compromise on these facilities, the answering of student questions in relation to their courses is essential.

The present technological opportunities that empower eLearning, for example are eLectures, e-books, netbooks, iPods and e-addressing. It recognizes and examines an educator’s perspective and the advancements on two essential aspects: the method for introducing and arranging learning information. With that, paramount attributes concerning the measure of intelligence of engineering or conceivable joint effort potential outcomes are overlooked. However, the emergence, development, improvement and across the board utilization of new correspondence and coordinated efforts are seen in Web 2.0.

It is more imperative that e-learning get to be more social and intelligent for learners. One can consider approachable web innovation like podcasting that additionally offers
learners the choice to eagerly make and improve sound recorded files independent from anyone else. An alternate illustration is a propelled web conferencing framework that utilized as synchronous online coordination and collaboration instrument by means of the web, progressively is constantly executed in elearning environments. There is no face to face interaction between students and other stakeholders like the academic staff, administrative staff, and registrar office staff. Coordination and communication will be through Web Services available on VU web server. To achieve objectives, everyone in E-Learning System will take the support of Web documents and Web services.

1.3 Research Gap

The topic of this dissertation lies at the intersection of the three emerging research fields: eLearning Management System, Multi-agent System and Question Answering System integration as shown in Figure 1.1. When analyzing Learning Management System (LMS) and agents based approaches for the question answering system, the main focus is made on the applications for e-Learning to emphasize their relevance to this dissertation. The field of multi-agent is reviewed from the point of integration and the sources of communication available for interactions to facilitate the students by the developed algorithm to more convenient in eLearning paradigm.

![Figure 1.1: Research Gap Identification for this Dissertation](image-url)
1.4 Approach of the Study

In this study, a conceptual framework for Learning Management System has been introduced. This framework describes the ways to model the learner and content, and also to find the best match between the learner question to the answer content. The present study is conducted to classify the delay in the existing system. The delay is further motivated for the automated question answering system. The agents are used in this framework to find the best match between the learner question and communications for the answer processing. Based on this framework, we designed and developed a multi-agent system. There are different agents each having specific roles in the system. JADE [19] framework was chosen to implement these software agents. Gaia [20] was chosen as an agent-oriented software engineering technique to analyze and design the multi-agent system. In literature, there are many questions answering systems providing with powerful features. However, it is not possible to integrate most of these systems with the existing learning management systems. Most of them were developed to function as standalone system. There are lots of learning management systems used in practice and it might be very effective to plug in features to these already existing and widely used learning management systems. In this research, a multi-agent base QA system was designed to be used with any learning management system. To achieve this interoperability, a communication protocol was defined to establish communication between the learning management systems and the Multi-agent Object Oriented Question Answering system, namely MOQA. The system was integrated into an open source learning management system, called OLAT\(^1\) [21].

1.5 Problem Statement

In order to minimize the gap between the student and the faculty member, it is important that student questions are responded to as soon as possible, and this may involve hiring new resources in every department, potentially in extensive numbers. It is well-known fact that by not increasing the human resources in accordance with requirements, the quality of education and productivity of the students will suffer. Administrators in eLearning institutions have launched an application that provides live audio interaction; it is installed on the PCs of every faculty member in order to minimize the gap between eLearning students and instructors. This is currently one of the most appropriate and

\(^1\)http://www.olat.org/website/en/html/index.html
feasible solutions, but it is costly and will only serve the campus-based students for two or more hours a day. This does not support job holders or home-based students. Another concern is that if the students get a reply through the application, their participation through MDBs will be reduced.

With the evolution and emergence of new web technologies, it is needed to review the system in order to bridge the gap in eLearning system and OER, then try to strengthen the web services and web documents for information retrieval. The machine must have to empower, strengthen, and semantically enrich documents. The agent can inform machines and computers what to ask and how to reply on a daily basis. In order to do so, we will have to convert the existing web documents and web services to semantically marked-up web documents. Let us consider a conceptual scenario comprising LMS, in which a student takes an online lecture course as shown in Figure 1.2. Suppose the student faces some sort of problem, confusion, or conflict in understanding any particular concept.

- What should be the next step?
- How will students minimize their confusion?
- What kind of service will be provided by the university in this regard?
• If the student posts the question on MDBs, then they must wait for a response. They may be able to get a response after few hours or perhaps a day, during that time the student remains unable to proceed, confused, and disappointed.

This leads to a decrease in education quality and ultimately student participation in MDBs may suffer. For our investigation, deciding about a suitable examination approach is essential, since it will help us reach specification with respect to our concerned question. Moreover, the determination of the investigation technique will help us to diagram the explanations and issues relating to our question. Delays in answering are complicated and complex, and it would be impossible to select an examination method that is within the limitations of what we can perform. A qualitative examination method offers results that are typically rich and distinct, demonstrating an enormous amount of arrangements. A qualitative investigation strategy has the ability to instruct us on how people feel and what they acknowledge.

This research on QA will be based on the breakdown of the standard components to improve the relevant activities. Google is good in case of response size, the options available and low time delay, but with minimal answer semantics and student’s understanding. Search on Google is performed with the help of keywords and response is generated with the ranking of documents. While the proposed system will be a smart intelligent system so the search will be in terms of semantic, not the keywords. Email is good in response size, answer semantics, and student’s understanding but time delay is very high and options available are really few. The Moderated discussion board (MDB) is the way used by Virtual University of Pakistan which allows students to post questions and instructor replies them one by one as shown in figure 1.2.

**Concept Understanding:** If learner faces some sort of problem/confusion/conflict in understanding any particular concept. What will be the next step? How he/she will minimize this confusion?

**Service Delay:** Service delay must be occurred and reduced the learning curve.

**Existing Moderated Discussion Board:** What kind of services are provided by universities in this respect? Exactly, he/she must paste the question at Moderated Discussion Board. The duration of getting an answer is unpredicted and without time limit. Coursera.com offers no direct question to the teacher.
**Student’s Perspective:** What will happen with the students? The students remain stuck, confused and disappointed. This leads to a decrease in Quality of Education and ultimately student’s participation over MDB will be at risk.

### 1.5.1 Research Questions

The research questions are a mandatory part of the thesis which are used to clearly understand the research problem. The research is conducted in the thesis aimed to provide the question-answering integration in the eLearning framework. This integration provides the most effective interaction between students and teachers to enhance the learning curve. This research is focused on answer the main question to elaborate the students learning.

The success of eLearning depends upon satisfaction of students. Satisfaction does not only depend on good lectures but also response to the questions. In class room based learning, teacher answers the questions instantly but in eLearning this is a problem which is required to be addressed. Discussion forums in Learning Management System (LMS) have been failed to satisfy student questions. Answers mechanism is dependent on availability of teacher. The need of automatic question answering system is apparent. This research contributes to an integrated MAQAS with LMS to address this problem.

- What is the current state of delay in answering the student questions in eLearning paradigm?

- What are the measurements of the delay faced by the students in answering the question in the eLearning environment?

- What are the possible solutions to address the problem of delay in question answering to satisfy the students in the eLearning system?

### 1.6 Aims and Objectives of the Study

This research is offering new innovations for better and upgraded engineering based frameworks for the eLearning plan. It is expected to lessen the reaction time of the student’s questions and offer them an impromptu inquiry answer session through the electronic e-studying stage so the collaboration of learners could be improved to accelerate their studying process. The delay figure in the asynchronous e-studying environment
could be killed via mechanizing the learner question reaction framework. Emulating targets are distinguished:

- To develop an integrated framework for the collaboration in the eLearning system between learner and educators.

- To reduce in response time for the students and minimize the delay in answering in eLearning System.

- To facilitate the students and also assist educators in the eLearning system as well as informal learning system.

- To implement for providing the more relative and close information to the students instead of a huge number of pages.

- To develop framework that will also be helpful for providing the suggestion against the questions and problems of the learner.

1.7 Significance of Proposed Study

Electronic learning becomes an industry and a research stream. The customers of an eLearning industry are the learners having different background and personalities. The ultimate goal of eLearning is to improve the learner’s learning and performance levels during online learning. To reach this goal, there is a need for studies not only to develop infrastructures and to deliver information online but also to improve learner’s learning and performance. The secondary purpose is to construct a conceptual framework dening learner modeling, content modeling enhances the learning outcomes. Student’s interaction in order to get answers in eLearning has different mediums and proposed intelligent system will be an addition to the mediums. The Moderated Discussion Bard (MDB) is the way used by Virtual University of Pakistan which allows students to post questions and instructor replies them one by one. The proposed system will try to minimize the time delay and optimized the solution of the learner answers more semantically. The learning system must be exible and suitable for any particular kind of students. One of the basic requirements of education in the twenty-first century is to prepare students for participation in a knowledge-based economy. Therefore, the online learning environments are required to respond to students according to their knowledge levels and to provide a different form of knowledge presentation for different knowledge levels. The
learning materials are needed to be presented according to the student’s preference and knowledge level, which makes the study processes more interesting and effective. Online learning environments put much work on the instructor’s shoulders. The instructors become less focused on the learning process because the management and administration of learning setting take too much time. Some examples of faculty barriers to distance learning are summarized as follows:

- Lack of support by the faculty. Faculty roles must change the most in administering distance learning programs.
- Faculty needs to change teaching styles to that of a mentor, tutor, and facilitator when necessary.
- Meet the needs of distance students without face-to-face contact.
- Change the course content to accommodate diverse student needs and expectations.
- Lack the basic skills or hardware to fully participate in distance education.
- The difficulty in giving immediate feedback.
- The difficulty in assessing student performance.

To support teaching and to facilitate learning, learning management system must provide individualized help just as a human tutor would. The responsibilities of faculty and students in distance learning environments are more than in the traditional learning environments. Until recently, a major requirement of the existing learning management system was to support both instructor and learners and to be ease of use. However, this seems to be no longer the main concern anymore. There is a need for smart learning environments offering personal services with capabilities to learn, reason, have autonomy, and be totally dynamic. The adaptive technologies might diminish the barriers to learning in distance education. This study is an attempt to design and develop a learning system that enriches online learning management system with adaptive instruction using agent technology.
1.8 Scio-Economics Benefits

Proposed solution may be implemented in other research areas that require keen focus in an agent-based system like Biology, Economics, Sociology, Medical etc.

- It will be cost effective
- It provides the faster delivery content
- More effective learning by the improving the response time
- Lower environmental impact
- Learners can go at their own pace, not at class fellows.
- Time in classrooms can be spent on questions / topics introduced by other delegates that are irrelevant to the needs of the individual learner
- There is less social interaction time.
- It takes less time to start and wind up a learning session
- Less travel time to and from a training event
- Learners learn what they need to learn, they can skip elements of a program they don’t need

1.9 Research Contributions

The field of QA system is an active area of research and a number of online QA systems are available. The research contributions on agent-based QA system (MAQAS) presented in this article include:

- A tailored agent-based eLearning framework.
- An eLearning model that automatically extracts the answers to the student’s question.
- A dynamic model for hierarchically clustering sentences as answers and a new sentence ranking function.
• A new answer presentation model in which answers, rather than document lists, are organized by question-orientated keywords.

• Finally, MAQAS is an online system that attempts to automatically answer complex questions and will narrow down the questions that a QA system is unable to answer.

1.10 Thesis Road Map and Conventions

Chapter 2 provides the review of literature about learning management system, information retrieval, information extraction, agents, and question answering system.

Chapter 3 presents the basic conceptual framework developed in this study for the delay assessment and measurement. The proposed method covers the student’s questions modeling and the learner-content matching strategies. The framework also provides the mechanism to assessment the need of an automated question answering system for the eLearning paradigm.

Chapter 4 contains the implementation details of the Multimedia and Agent-based Question Answering Framework for the eLearning System, MAQAS. The agents with interactions among the agents and Ontologies developed to establish agent communications are explained in this chapter. This chapter focuses mostly on the technical architecture of the system. This chapter also provides screen shots of integration of the system into the learning system integrations.

Chapter 5 provides the discussion of the question answering system and their performance analysis. The proposed system comparisons with the existing question answering system on a different type of question categories and their types like factoid, definition, and descriptive questions. This chapter also includes the image based result analysis.

Chapter 6 presents the conclusion and possibilities for the further directions in research.
Chapter 2

LITERATURE REVIEW

This chapter discusses the background concepts and services for the eLearning paradigm. It also introduces the agent technology, agent types, intelligent agents and multi-agent systems. The application domain for the agent types and there uses for the eLearning paradigm. Web semantics utilization and web semantic services for the eLearning paradigm. Moreover, use of web services for the question answering system. The multi-agent system is used as fundamental in eLearning platform for the question answering system to facilitate students.

2.1 Background

eLearning is designed for instructors, learning technologists, and people with a general interest in education. Learners who want to deepen their knowledge use eLearning courses and training from international online universities. In contrast to traditional classroom learning, eLearning is defined as knowledge collaboration using electronic media. It comprises spoken and printed word, as well as static pictures or videos, delivered with the help of electronic technologies. Content delivery and management for eLearning rely on electronic resources (eResources) and Information Technology (IT). A number of internationally ranked universities have started distance learning courses, including Stanford University, University of New Mexico, University of Houston, Duke University, University of London. Coursera\textsuperscript{2}, Edx\textsuperscript{3}, Khan Academy\textsuperscript{4}, CourseSites\textsuperscript{5}, and many more collaborative institutions are also offering eLearning courses. CourseSites has

\textsuperscript{2}https://www.coursera.org
\textsuperscript{3}https://www.edx.org
\textsuperscript{4}https://www.khanacademy.org
\textsuperscript{5}https://www.coursesites.com
reshaped the traditional educational concept to eLearning by providing online, free of cost, and education for all students. CourseSites learning resource has 74,200 instructors belong to 12,573 institutions in the 159 countries offering various courses. Moreover, another well-known eLearning portal is Coursera. At Coursera, almost 1,434 courses are offered in various languages. These courses cover business administration, computer science, arts and humanities, physical sciences, life science, etc. More than one thousand courses are offered in English and many other languages are also represented.

2.1.1 Factors Increasing eLearning Enrollments

Electronic Learning courses are offered by internationally recognized faculties and universities. Table 2.1 indicates the factors behind increases in the percentage of eLearning enrollments.

<table>
<thead>
<tr>
<th>Reason</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic downturn percentage</td>
<td>37%</td>
<td>22%</td>
<td>18%</td>
<td>40%</td>
</tr>
<tr>
<td>Typical distance education growth percentage</td>
<td>39%</td>
<td>28%</td>
<td>28%</td>
<td>36%</td>
</tr>
<tr>
<td>New enrollment initiative percentage</td>
<td>12%</td>
<td>14%</td>
<td>9%</td>
<td>17%</td>
</tr>
<tr>
<td>Unknown percentage</td>
<td>5%</td>
<td>7%</td>
<td>11%</td>
<td>18%</td>
</tr>
<tr>
<td>Other percentage</td>
<td>7%</td>
<td>13%</td>
<td>21%</td>
<td>17%</td>
</tr>
</tbody>
</table>

2.2 Course Offered for the eLearning Paradigm

The list of courses offered, both basic and advanced, continues to increase. In addition to being offered in English and other languages, such as Chinese, French, Turkish, Russian, Spanish, many courses are offered with subtitles in this eLearning paradigm.

Coursera eLearning website data shows the courses offered with subtitles. These are 1,434 courses offered world widely. For the list of disciplines, a Preto chart is plotted. Discipline distributions for the eLearning course data are shown in Figure 2.1. This data is in descending order of frequency with an accumulative line on the secondary axis as a percentage of the total, which describes the boost in eLearning and other distance learning paradigms. There are 1,408 courses offered through English, and 1,434 courses are offered with subtitles in other languages. A tree map is drawn for the courses currently available. Figure 2.2 shows the number of open courses, as well as other categories of self-paced and pre-enroll courses.

These online courses have different durations, from 1 to more than 3 months. There are 604 open courses, 149 self-paced courses, and 119 pre-enroll courses. There are 22
courses of 1 to two months duration, five courses lasting 2 to 3 months, and two courses taking more than 3 months. In Pakistan, many institutions established their set up for eLearning as like many other world-class universities. This list is including the Virtual University of Pakistan, Virtual COMSATS, GCU Faisalabad Distance Learning Program UET, NIMLS Online Learning. The student learning curve decreases drastically in the eLearning paradigm due to the absence of tutors to respond to student questions.

2.2.1 Top 10 eLearning Websites

Top 10 links of the eLearning websites are shown in Table 2.2. These top 10 links are selected from the top most 40 web resources as proof of concept. These resources are used over the world for the learning and training purposes.
Table 2.2: Top 10 eLearning Websites and Resources for the Students

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Website</th>
<th>URL</th>
<th>Detail</th>
<th>Source / Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ALISON</td>
<td>alison.com</td>
<td>60 million lessons</td>
<td>1.2 million unique visitors/ month</td>
</tr>
<tr>
<td>2</td>
<td>COURSERA</td>
<td><a href="http://www.coursera.org">www.coursera.org</a></td>
<td>Educational website that works with universities</td>
<td>Free to use and Learn from over 542 courses.</td>
</tr>
<tr>
<td>3</td>
<td>UDACITY</td>
<td><a href="http://www.udacity.com">www.udacity.com</a></td>
<td>Advance the education and career through project-based online classes</td>
<td>Focused on computer, data science, and mathematics.</td>
</tr>
<tr>
<td>4</td>
<td>MIT Open CourseWare</td>
<td>ocw.mit.edu/index.htm</td>
<td>Free access MIT courses</td>
<td>expect from MIT.</td>
</tr>
<tr>
<td>5</td>
<td>Khan Academy</td>
<td><a href="http://www.khanacademy.org">www.khanacademy.org</a></td>
<td>Thousands of micro-lectures on topics</td>
<td>from history and medicine to chemistry and computer science.</td>
</tr>
<tr>
<td>7</td>
<td>Maths &amp; Science</td>
<td>mathsandscience.com</td>
<td>Courses</td>
<td>tests and learning materials about mathematics and science</td>
</tr>
<tr>
<td>8</td>
<td>edX.org</td>
<td><a href="http://www.edx.org">www.edx.org</a></td>
<td>Free courses designed specifically for interactive study via the web,</td>
<td>MIT, Harvard, Barkley, Georgetown, Boston University of Washington, Kyoto &amp; much more.</td>
</tr>
<tr>
<td>9</td>
<td>iTunes U</td>
<td>itunesu.itunes.apple.com</td>
<td>Apple’s free app that gives students mobile access to many courses</td>
<td>free video courses, books, presentations and audio lectures.</td>
</tr>
<tr>
<td>10</td>
<td>Codeacademy</td>
<td><a href="http://www.codecademy.com">www.codecademy.com</a></td>
<td>An easy way to learn how to code.</td>
<td>interactive, fun and you can do it with your friends.</td>
</tr>
</tbody>
</table>
2.2.2 Learning Resources

Learning resources like eBooks, journals, online journals, and even general correspondence on a theme by a group can be utilized as a critical source of data. The recovery of applicable data from resources with huge web documents requires substantial time and effort. In light of Open Educational Resources (OER), Massive Online Open Courses (MOOCs) stand out as being among the most flexible approaches offering access to quality training, particularly for those living in remote or troubled regions [27]. MOOCs could be used for casual eMentoring, a sub-assignment of eLearning. AMAS is suitable for applications requiring inference, judgment, and decision-making mechanisms, process learning, and the need to choose between alternatives in multiple-option solutions.

2.2.3 Online Learning

Learning resources might be as ebooks, ejournals, masters’ online journals, and even general correspondence over a theme around the group could be utilized as a critical source of data. Recovery of applicable data from assets having huge voluminous Web and their association for later utilize is a tedious and lengthy work [15].

2.3 Web Semantics

Web 1.0 has limitations, which have been pointed out and ultimately web 2.0 continuously replacing and reshaping the web 1.0. We have introduced the Semantic Web and main focus was on its core assumptions and solutions. Finally, Semantic Web 2.0 is presented, which links the Semantic Web platform into existing Web 2.0 features.

2.3.1 The Current Web

Web pages that run over the internet containing multimedia text images are interlinked hypertext documents is the invention of the Sir Tim Berners-Lee that are navigated by hyperlink with the web browsers [63]. Each page written in HTML has its own URL. Web sites are connected each other through hyperlinks. Although there is a lot of information exists that is not machine readable as a human can understand words while particulars for the machine does not exhibit. People looking on the web then why the machine should understand the web contents, this can be a question mark. A man who is fond of music especially rock and ska once hear reggae song by Bob Marley that was much liked by him. To learn about that, he joined reggae music fan community.
Chapter 2. Literature Review

However to connect with that community or to find out the fans of Bob Marley he needs a precise statement to be written down on the web to get required information. Although its computer task to find out answers to our question, it would not be accurate if the given information is not comprehensive and precise, this example shows that appropriate resource definition is a most important factor to get the desired output. There must be some mechanism for the machine to distinguish a different piece of the information.

2.3.2 Web 2.0

An interactive space to connect people using social software applications where they can meet, collaborate and share is called the 2nd generation of the internet based services including blogs and communication media like Skype, Wikipedia is Web 2.0. RIA, XML, CSS, XHTML are mostly being used internet applications that make the URL clean and meaningful are derived from Web 2.0 as shown in Table 2.3.

<table>
<thead>
<tr>
<th>Table 2.3: New trends in the Web Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Web 1.0</strong></td>
</tr>
<tr>
<td>Platform</td>
</tr>
<tr>
<td>Web pages</td>
</tr>
<tr>
<td>Portals</td>
</tr>
<tr>
<td>Encyclopedia</td>
</tr>
<tr>
<td>Arrangement</td>
</tr>
</tbody>
</table>

2.3.3 The Semantic Web

The eLearning requirements realized that semantic web is considered as a promising technology. Sir Tim Berners-Lee said Semantic web facilitate the users to move on pages via software agent for getting required contents of the web. The Semantic Web supports the word semantics by building a new W3 architecture that shows further clearance of the required contents as shown in Figure 2.3. Through Semantic Web information can be shared, searched, browse as it works like a universal shared medium [62, 17]. Flexibility in the semantic web where different analysis on gathered data carried out is a significant feature of the cyberspace [64]. Set or modules are used to enhance the vocabulary [7].

2.3.4 Modeling Ontology

A number of approaches were considered by the World Wide Web Consortium (W3C); these are the Web Ontology Language (OWL-S), Web Service Description Language
WSDL-S), Web Service Modeling Ontology (WSMO), Web Service Execution Environment (WSMX), and Semantic Annotations Web Service Description Language (SAWSDL) [28]. OWL-S makes it possible for agents to achieve dynamic discovery, composition, invocation, and monitoring of services. OWL-S has three parts: the first is the “profile”, which is used to describe the web service, as well as the discovery information; the second part is the “process model”, which deals the service operations; and the final part is the “groundlings”, which illustrates interoperation through messages within the service [29]. A WSMO approach also has different parts: the ontology and the modeling language and execution environment [30]. It provides the same services as OWL-S, i.e., annotates input, output, and methods for the description to achieve dynamic discovery, composition, invocation, and monitoring of services. WSDL-S introduces a way in which web service descriptions like input, output, effects, methods, and category information can be semantically annotated using the ontology. Another similar approach to WSDL-S is SAWSDL, which defines some WSDL extension attributes along with an XML schema to semantically enrich WSDL components [31]. Some integrated frameworks have been developed to achieve SWS functionalities. WSMX provides an integration execution environment for the dynamic discovery, mediation, interoperation, and invocation of SWS.
IRS is another example in this context. In the context of the relationship of intelligent agents (an autonomous piece of code capable of extracting the semantics of data) with semantic web technology (in which services can be semantically enriched). It has been concluded that to understand, process, and execute services effectively. An intelligent agent is needed to achieve outcomes at a high level. All of the above-discussed approaches to SWS technology depend on upon intelligent agents for dynamic discovery, composition, invocation, and monitoring of web services [32].

2.3.5 Ontology Development

Ontology development is an important part of web semantics. Before moving on to semantic web services (SWS), it is necessary to look at some general characteristics of web services, including their benefits and shortcomings [30]. A web service is hosted at some point on the Internet and can be accessed from anywhere by standard protocols. Web service standard protocols include Universal Description Discovery and Integration (UDDI), Web Services Description Language (WSDL), and Simple Object Access Protocol (SOAP) [28]. UDDI provides a way to find particular web services as well as helping the user to publish in a directory. WSDL describes a web service, its input, output, functions, method, and grounding details written in XML format. SOAP provides interoperability in a heterogeneous environment for message exchanging. This current technology is flawed in relation to heterogenic diversity. Current standards are unable to provide a facility for the dynamic discovery of a web service, its invocation, or its composition, meaning there is a need for automation. To achieve this requires a joint endeavor of the semantic web and web services. Intelligent agents are used in this way to process useful semantic information.

2.3.6 Semantic Web Services

Several projects have been carried out with the aim of combining the strengths of SWS’s and intelligent agent technology. It can be done in such a way that alleviates their respective weaknesses. An effective method for integrating these two is still required. Particularly one addressing the fact that web services provide a very basic level functionality, whereas agents provide a higher level functionality in communication, and integration with gateways [32], [33]. The integration was done in a manner that describes the invocation of the service using the ontology languages, The Defense Advanced Research Projects Agency (DARPA) uses DARPA Agent Markup Language (DAML) and
Ontology Infrastructure Language (OIL), with the agents to invoke the service [38], [58]. Another study was carried out for a Small Web Format (SWF) Project, in which agents perform various activities related to the owner and resolve issues accordingly with the help of web services and computational resources [52]. An IST SPICE project illustrates the dynamic discovery, composition, and invocation of web services with the flaw that it was strictly coupled with WSMO [29] [55].

Moreno [34, 35] and his colleagues recommend that energetic pedagogical negotiators may help students develop a responsive connection with the agent. The agent is then assisting the students with their choice of learning conditions. Initially, there are cognitive benefits of using the agent system, but there are also some disadvantages of using this approach. This method cannot be used on a large-scale due to its lack of suitable communication protocols, and a adequate solution must be found. This approach also has a lack of dynamic aspects. Another disadvantage of using agents is the non-effective use of semantics during communication. In the context of associations, an intelligent agent is an autonomous piece of code able to extract data semantics. In semantic web technology, services can be semantically enriched. It has been concluded that to understand, process, and execute services effectively, an intelligent agent is required. All of the approaches to semantic web service technology discussed above depend on upon intelligent agents for dynamic discovery, composition, invocation, and monitoring of web services [16]. Deployment of a Multi-Agent System (MAS) framework will be helpful many fields such as medical, telephone, space science, and business expert systems.

The general characteristics of a MAS, and the strengths of this system, which combine to achieve more beneficial results. A single isolated agent can be beneficial and accomplish work for the user; however, more often, numerous agents work in cooperation to form a MAS. Along with all of the benefits of an isolated agent, a MAS has numerous other benefits, including trustworthiness, concurrency, parallelism, scalability, and flexibility [12]. In a MAS, some issues of concern are coordination and the communication language. The two MAS languages are the Knowledge Question and Manipulation Language and the Foundations for Intelligent Physical Agents-Agents Communication Language. MAS’s can be deployed in medical, telephone, space science, and business fields, among others, and the INGENIAS programming methodology is the most suitable for use with them. In MAS’s, standardization regarding the interoperability of different agents in a heterogeneous environment is an open-ended question for larger organizations. There are also other research areas that require attention for agent-based systems;
these include the introduction of new techniques and methodologies, the creation of tools for MAS development, management, and integration, along with the development of its connection with other branches of knowledge, such as biology, economics, sociology. As the web expands, people want to find their required information immediately. To achieve this, the semantic web provides a way to add semantics to published data in such a way that machines can understand and process it in a human alike mode. Stoyanchev [13] introduced an information retrieval QA framework, and assessed the utilization of named entities as well as verbs and prepositional expressions, as accurate match phrases in a record recovery question. Kangavari et. al has demonstrated the least complex way to deal with enhancing the precisions of a QA framework. Martin et al. [14] appear to contemplate Relational Database Management System (RDBMS) usage for the IR that utilizes an OWL ontology with three interrelated sub-ontologies: the profile describing What a service does, the process model narrating How it works, and the grounding describing How it communicates. Paloma Moreda Hector Llorens [15] presents two recommendations for utilizing semantic data as a part of QA system, particularly in the answer extraction step. His point is to decide the change in the execution of current QA frameworks, particularly when managing basic questions. Aldao [16] discusses the configuration of element information based on a full-message recovery framework and record innovation research and examination, given some of the indexing code, so as to enhance the recovery precision.

Considering pedagogical improvements of this composition moreover, anticipate closeness issues with current advancements e.g. IEEE LOM recollecting the last target to perform the best interoperability[17]. Numerous LMS’s are deployed in different educational institutes with the purpose of providing learning facilities through web services, including MDBs, where student and instructor communicate via a board. Other services include online quizzes, assignments, and exams. It is evident that such applications have become more complex. The type of question has been changed in this system to cope with complex questions [36].

2.4 Agent Technology

The agent is a piece of code that exhibits like a human being and takes decision against some inputs. The agents are used in different types of applications that have complexity ranges; small, medium, and large. The small-scale uses of the agent technology are in email filtering. The large-scale usage of the agent technology is in Air Traffic Control
system. In general Agents either a physical or virtual entity that can possess the any of
the following definitions:

- Capable of acting in environment
- Make possible communications with other agents,
- Hold the owner’s resources
- Agent can perceive from the environment
- Offers different service and skills
- Capable of the reproduction
- Whose aim to produce satisfactory results by using the resources and skills on the
owner’s behalf, to its possession, perception representation, and communications.

2.5 Agents Types

The agents have different types according to their functionalities. The agent types for
common use are following:

2.5.1 Mobile Agents

The mobile agent’s main feature is mobility, which can travel from one host to other to
perform the assigned tasks on the peripheral hosts [18]. Mobile agent frameworks are
rare because the high level of interest is required to allow foreign agents on data servers.
The well-known example is Sony’s Magic Link PDA agent which manages the user’s
phone, fax, email and pager along with linking the user to Tele-Script that provides
services for communication and messages [43].

2.5.2 Collaborative Agents

Collaborative agent’s are mostly used for information sharing. These agents exhibit
autonomous behavior and cooperate with other agents as well. These agents perfor-
m specific tasks for their master in the open and closed time constraints multi-agent
paradigm [18]. These agents are used in order to solve large-scale problems and the can
negotiate with another agent to reach an agreement. The also used in the reasoning
purposes. ADEPT is a well-known example of the collaborative agents[43].
2.5.3 Reactive Agents

Reactive agents are also called as autonomous agents and they react with the environment stimulus responsive manner [43]. They are commonly used in embedded fashion as these are most robust, powerful and fault tolerance than other agent system.

2.5.4 Interface Agents

The interface agents are used as a personal assistant and for the collaboration with end users. Interface agents only communicate with users instead of collaborative agents, they reduce the work for the developer and end user [44]. They take time for learning and adapting according to the preferences and habit of the user. Calendar Agent is a well-known example.

2.5.5 Hybrid agents

Hybrid agents are composed of more than one above discussed type agent’s characteristics. They are used to provide the maximum utilization of the resources and benefits. These agents are a combination of the reactive and premeditated properties.

2.5.6 Internet / Information Agents

The internet agent can also be called as information agents. These are used for the manipulation, management and collation of the information from the intranet. These agents fulfill the information management problems and sometimes also act as mobile agents. Jasper is a well-known example of the information agents that can store, retrieve, summarize and communication with another agent over the Internet [43].

2.5.7 Service Agents

Service agents are responsible for facilitating the other systems, agents, and users low-level services [44]. These agents are universally trusted agents and commonly provide the reporting, logging, and director management information.

2.5.8 Intelligent Agents

An intelligent agent is a piece of code that executes and follows the way like a human by the Nwan and Ndumu [43]. An agent as a module may be hardware or software which can perform tasks on behalf of the owner or user. Some general characteristics of intelligent agents are: work in a constant and continuous way, have an ability of decision making upon the given information according to the current situation, a property that
urges an agent to maximize its effort in order to achieve desired output, interoperability and most importantly a learning process from context and by user input [2].

As the web grows extensively every day and people want to find out their specific information without delay. So in order to achieve this cause, the semantic web came in front, provided a way to add semantics to published data in such a way that machines could be understood and processed in a human-like way. Ontology development is an important technology in this approach [7]. We will discuss some general characteristics of agents, semantic web, semantic web services, eLearning case study scenario for conceptual level as well as for analysis level, the motivation of this work and benefits of combining these technologies. Another section contains the proposed system and its working with respect to e-learning by combining SWS technology and multi-agent technology. The integration was done in a manner that describes the invocation of the service using the ontology languages, Defense Advanced Research Projects Agency (DARPA) Agent Markup Language (DAML) and Ontology Infrastructure Language (OIL), with the agents are used to invoke the service.

The Intelligent Agent describes as the program code than used on world wide web to accomplish tasks of accepting, delivering, retrieving information and auto complete the repetitive the tasks. An agent is an entity that performs for others with the properties of competency, delegacy, and amenability. Acting autonomously is called delegacy and competency comprise of the communication proficiency. Adapting the behaviors for the optimized performance in a dynamic environment is known as amenability. Intelligent Software Agent (ISA) that uses the artificial intelligence to accomplish the user’s tasks actively. Multi-Agents Systems shared and reduced the workload of the owner’s by a delegation of the intelligent software agent.

2.6 Multi-Agent System

This section discusses the general characteristics of a MAS, and the strengths of this system, which combine to achieve more beneficial results. A single isolated agent can be beneficial and accomplish work for the user; however, more often, numerous agents work in cooperation to form a MAS. Along with all of the benefits of an isolated agent, a MAS has numerous other benefits, including trustworthiness, concurrency, parallelism, scalability, and flexibility [37]. In a MAS, some issues of concern are coordination and the communication language.
The two MAS languages are the Knowledge Question and Manipulation Language and the Foundations for Intelligent Physical Agents-Agents Communication Language. MAS's can be deployed in medical, telephone, space science, and business fields, among others, and the INGENIAS programming methodology is the most suitable for use with them. In MAS's, standardization regarding the inter-operation of different agents in a heterogeneous environment is an open-ended question for larger organizations. There are also other research areas that require attention for agent-based systems; these include the introduction of new techniques and methodologies, the creation of tools for MAS development, management, and integration, along with the development of its connection with other branches of knowledge, such as biology, economics, sociology. As the web expands, people want to find their required information immediately. To achieve this, the semantic web provides a way to add semantics to published data in such a way that machines can understand and process it in a human alike mode.

2.7 Agent Application Domains

Agents are mostly being used either to find out the solution of such problem that cannot be solved by the existing technologies or solutions are too expensive or to mold the already existing solutions in an efficient, easier, cheaper and natural way. Main domains where agents used are industrial, commercial, medical or entertainment [45]. In industrial agents used in control process manufacturing or air traffic control. Information management, business process management and in electronic commerce agents are being used in the commercial side [46, 47]. ADEPT agents used in the application domain, patient health can be monitored by the agents in medical side. Agents utilized in animated movies or characters, interactive games and theaters [54].

2.8 Agent Development Environments

A number of Agent development environment are used for the agent development as JADE [19], FIPA-OS [48], ABLE [49], JATLite [72], Aglets [70], AgentBuilder [69].

2.8.1 JAVA Agent Development Environment (JADE)

JADE is FIPA-Compliance Java Development environment, which is developed in Torino, Italy. The multi-agent system adopted the JADE environment for the creation and management of the agents [50].
2.8.2 FIPA-OS

FIPA-OS agent development environment is designed for the Nortel Networks for the communication language standardization [23]. IEEE Computer Society adopted FIPA officially in its eleventh standard committee 2005. Now FIPA is a standard organization for the promotion of the Agent Technology.

2.8.3 Agent Building Ad Learning Environment (ABLE)

ABLE is also a Java based framework with the toolkit, and library for the intelligent agent development by using the reason and machine learning [49]. It provides the AbleBeans which is JavaBean components with several interconnection methods for development of software agents.

2.8.4 AgentBuilder

AgentBuilder is integrated toolkit that is used for the speedy development of intelligent agent and software based on agents [51]. The Reticular Systems Inc. is developed AgentBuilder and categorized in two products. AgentBuilder Lite is used for the single agent and a small application. The other one is AgentBuilder Pro which used for the Multi-Agent system.

2.8.5 Aglets

Aglets is also another project of the IBM, which is Java based and facilitate the java object that can transfer from host to host [52]. The aglets executing on one host and may transfer at any time to another host as per owner’s requirements.

2.8.6 Java Agent Template Lite (JATLite)

JATLite is developed by the Stanford University and used for the Multi-Agent system and applications. It includes light-weight Java Libraries [53]. JATLite allows users to develop quick software agents for the deployment and communication over the internet. Java Agent Development Environment is used for the development of the agents. In this research, the JAVA Agent development environment is used for the creating the multi-gene system. JADE provides the set of packages for the development and deployment of the distributed agents. The JADE service for the agents has transported protocols, agent naming service, yellow pages. JADE provide the more user-friendly graphical Unser Interface for the better control and monitoring of the agents, which are running at JADE platform. The requirement for the question answering system for the student
is a Multi-agent system, hence JADE opts for the Agent development platform for this system.

2.9 Question Answering

In eLearning paradigm question, answering is a mandatory part. In the existing eLearning frameworks, automatic question answering facility for the student is missing. From the last 20 years, a number of query answering systems have been produced utilizing the new ideas and some different techniques. Hyo Jung and the others [54] described the classification of Chinese questions on the base of rules of the mining association which except the word and the bi-gram from queries as great features. The Jieyu and Kepei [55, 56] conferred a Chinese Question Answering System with the classification of question which utilizes Word, Part of Speech (POS) tagging, the named entity (NE) and linguistics as an exemplary element for classifying the respective question. Santosh Ray [57] explained about a portion of the current methodologies for query classification and proposed another strategy through the utilization of the WordNet. Svetlana [58] conferred a procedure of the document retrieval for a query answering system, and adopted the process of utilization of named elements, nouns, verbs, and phrases as accurate match phrases in a query of document retrieval. While Kangavari and the other [59] displayed least difficult way to deal with enhancing the precision of a query answering system. The Chiyoung and the others [60] explained in their performance review and studied that the implementation of RDBMS utilizing the inverted index method must surpass the implementations of Information Retrieval.

Paloma and the others [61] explained about the two recommendations for utilizing linguistics data as a part of Query Answering system, particularly in the step of answers extraction. Its purpose is to find out the enhancement in performance of existing Query Answering Systems, particularly when these QA systems give the answer of those questions which contained the usual noun. Liang Yunjuan and the others [62] explained about the design of text retrieval system which depends on the learning based dynamically, inverted index technique and analysis, provided various indexing code for enhancing the precision of text retrieval.

Numerous architectures have also been presented for making the different Question Answering Systems. Mones and the others [63] have given a paradigm taking into account the interpretation of query statement. Moreover, they also gave the model
about the translation of the document into a format of human-computer readable which is a tiny bit matured and costly. Mohammad Reza and the others [59] have additionally proposed a design in 2008 utilizing the dynamic examples and syntactic relations among nouns, verbs, keywords and the phrases words. These both structures may perform fine, yet these both architectures are extremely complicated containing many modules which are hard to develop. Some frameworks have also been designed by Harabagiu [64] and Hrihan and the others [65], these frameworks look like to be simple and are effortless to implement by architecture, yet the issue is that the phases of their working components are not obvious to comprehend.

Shen and Lapata [66] determined the involvement of semantic roles to factoid type of query answering and demonstrated promising results. Pizzato and Molla [67] proposed a query forecast language model is giving rich data and accomplished better pace and precision. Although their work is differentiated from others because they consider numerous fields while the others consider only one field is presenting to semantic roles. Ferrucciet [68] displayed IBM Watson adopting a hybrid approach between Natural Language Processing and Information Retrieval, and propelled the query answering work to another level. Fader [69] proposed an idea to extract the summary through supervised learning approach utilizing a seed lexicon. For this purpose, they used the “Perceptron Learning Algorithm”.

A learning process in this proposed research work is comparative; in any case, it is differentiated in a way that we learn weights for individual fields rather than dictionaries or lexicons. Yih and others [70, 71] presented a semantic parsing system for open domain query answering, which utilized complication neural systems for measuring likenesses between crumbled entities. Weston and others [72] introduced the Memory Networks models build to retain information about known articles and actors. Several QA systems used in the literature have been discussed in next subsections.

2.9.1 START System

The abbreviate of START is ”SynTactic Analysis utilizing Reversible Transformations” that is the question answering system for the human natural language. START is the first Web-based query answering system in the world that has been on-line and continuously working since December 1993 [59]. This was designed by Boris Katz in AI Laboratory at MIT’s and a key system called ”natural language annotation”. This was utilized as a part of a pre-compiled learning base to give the answers to questions [48].
2.9.2 Answer Bus

Answer Bus is a question answering system for the open-domain. Answer Bus is developed on the base of sentence level Web data retrieval. It takes natural language queries from users in multi languages and concentrates conceivable answers from the Web. Answer Bus can also give answers to the users’ questions in little time.

2.9.3 Language Computer Corporation (LCC)

LCC’s Question Answering System [73] was created by the company of Language Computer Corporation (LCC). This system was so good in TREC-2002 Question Answering, TREC-2003 Question Answering and the TREC-2004 Question Answering [73-78]. Language Computer Corporation succeeded in light of joining the qualities of Information Extraction procedures with the huge representations of axiomatic facts and knowledge obtained from WordNet to defending answers which are extracted.

LCC used many sophisticated tools and in order to improve the performance of their system some other tools also have been used which are Named Entity Recognizer (NE), Lexical chained, Logic Design Transformer, Disambiguate of Word Sense, Syntactic Parser, and Logic Proven, and so on. So it obtained 70-80% precision for factoid queries at TREC 2003 and TREC 2004 [79]. Anyhow, these Question Answering frameworks are yet not prepared to give the accurate and precise answers to the queries. In current years, different web based completely automated Question Answering systems in the English language to give precise answers for respective queries have been produced which depends on various search engines [56]. It adopts a novel based architecture that consolidates data retrieval concepts with the processing of natural language, parsing of the query, answers extraction and gives voting to answers for selection. The system works for generating a particular search against the reformulated query, and then extracting the relevant answers from the paragraphs in the searched page depend on the information of categories of queries.

2.9.4 AnsweFinder

Greenwood in 2005 explores various novel techniques and methodologies for open domain query answering [80]. Some techniques are analyzed and researched are; manually and constructed query analyzers without human intervention, retrieval of document particularly for query answering, extraction of the syntactic answer, and extraction of answers automatically through the text patterns matching techniques. The previously
stated work describes approaches on Factoid and Definition Question Answering methods. The novel base methodologies and few techniques in the paper are merged to make two end-to-end query answering systems so that answers can be found quickly rapidly: AnswrFinder and Varro.

Varro constructs definitions for terms, for example, "what is a computer machine?" and "define Random Access Memory". Both systems permit users to discover answers to their queries utilizing web documents that are retrieved by Google. These two systems together show that the techniques adopted in the paper can be effectively used to give fast and efficient answers for open-domain question answering.

2.9.5 Question Answer System for the Bengali Language

The Bengali language is one of the mostly speaking languages in the world with more than 200 million speakers. Anyhow, having such a massive speaker base the language needs a significant number of the fundamental language processing tools and resources which are already available in different languages [58]. The inspiration of the researchers is that there are unknown initiative activities for an advanced Bengali Query Answering system and attempted to handle this opportunity and propose a fundamental Query Answering system for Bengal. At that point in 2010, researchers investigated the possibility of a cross-language Query Answering system where a user would pose a question in the Bengali language. This system would create an answer from written texts in another language excepting of Bengali language and translate that language back to the Bengali language for the user. However, they achieved the 53% correct results to translate queries from Bengali to other specific English language.

2.9.6 Chinese Question Answer System

Since retrieval of Chinese text has quite been created recently and there are several different particular qualities in the Chinese language. Hence the research in Chinese Question Answering systems using traditional language had been started afterward than that of Japan and other western countries. The developers have confronted the immense difficulty to make the functional Chinese Query Answering Systems that utilize syntactical information. Therefore, the methods of statistical retrieval and analysis of superficial language are essentially utilized in question answering systems. On the web based Query Answering Huang endeavored in 2004 to represent texts in the Chinese language by its different features, and attempted to convert the text information to Chinese language.
into ERE relation and afterward to give an answer of the respective query using entity relation model [59]. Summing up, even though a number of the literature on query answering discusses the decisions have been made for preferring certain retrieval strategies over others, so those choices are rarely explained by comparative tentative conclusion-s. Table 2.4 describes the comparison among the different techniques of different QA systems [65].

**Table 2.4: Comparison of Different Techniques among Question Answering Systems**

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Types QA Systems</th>
<th>Methods used in QA systems</th>
</tr>
</thead>
</table>
| 1       | Question Answering framework utilizing syntactic and semantic data. | Chunking and Tagging to sentences, 
Named Entity (NE) recognition, calculating distance semantically |
| 2       | A QA system depends on corroboration.                 | Predicting of Answer Type, Named Entities (NE) existence, Finding Acronyms                  |
| 3       | Question Answering System propped up by data extrac-tion. | Named Entities based question answering, Question Processing, Text matching and handling technique |
| 4       | Question Answering System for multi languages queries | Syntactic technique, Answering structural design                                           |
| 5       | Multi languages Question Answering System             | Tokenization, Part of Speech Tagging, Disambiguation of Word sense, Answer type detection, Keywords expansion |
| 6       | Question Answering System Multi-language, Spatial transient context understanding, Temporal requirement | Syntactic technique, Answering structural design                                           |
| 7       | Evaluation of the Inquire Question Answering System   | Reformulating process of Query, N-Gram mining, tiling, and filtering                       |
| 8       | Hybrid Question Answering System depends on retrieval of information and the validation of answers | Module of Pattern Creation, Axiom Producing Module, Indexing of Document, Processing of Document |

### 2.10 Summary of Literature Review

Many QA systems are available and are also accessible publically on the Web, some of the following online systems are START\(^6\), AskJeeves\(^7\), LCC-Web\(^8\), YodaQA Live\(^9\),

\(^6\)http://start.csail.mit.edu/index.php
\(^7\)http://uk.ask.com/
\(^8\)http://www.languagecomputer.com/demos/
\(^9\)http://live.ailao.eu/
AskHERMES\textsuperscript{10}, Answers\textsuperscript{11}, and Quora\textsuperscript{12}. The student questions are answered manually by the instructor. In the absence of the instructor, the student questions remain unanswered. The automatic reply to the student question in the existing system is summarized as shown in Table 2.5.

The automated question answering (QA) track, which has been one of the most popular tracks in TREC for many years, focuses on the task of providing automatic answers for human questions. The track primarily deals with factual questions, and the answers provided by participants are extracted from a corpus of News articles. While the task evolves to model increasingly realistic information needs, addressing question series, list questions, and even interactive feedback, a major limitation remains: the questions do not directly come from real users in real time. The LiveQA track revives and expands the QA track, focusing on live question answering for real-user questions this year \cite{85}. Real user questions, extracted from the stream of most recent questions submitted on the Yahoo Answers (YA) site that has not yet been answered by humans, will be sent to our systems. Then our system provides an answer in real time \cite{60}.

Quora is a fast-growing social QA site where users create and answer questions, and identify the best answers by up-votes and down-votes with crowd wisdom. The Web of Data has grown to contain billions of facts pertaining to a large variety of domains. While this rich data can be easily accessed by experts, it remains difficult to use for non-experts \cite{86}. The pre-implemented HAWK system, however, performs better on hybrid questions than on plain English questions. Systems like YODA QA Live, which do only provide answers without an SPARQL query cannot be analyzed sufficiently \cite{87}. AskMSR to a new question answering dataset created using historical TREC-QA questions, we find that the key step, query pattern generation is no longer required but instead, deeper NLP analysis on questions and snippets remains critical \cite{86, 87}.

\textsuperscript{10}http://www.askhermes.org/index2.html
\textsuperscript{11}http://www.answers.com/
\textsuperscript{12}https://www.quora.com
### Table 2.5: Summary of Literature Review

<table>
<thead>
<tr>
<th>QAS with Author</th>
<th>Year</th>
<th>Technique</th>
<th>Domain</th>
<th>API Used</th>
<th>Answer Processing Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>START, Katz</td>
<td>1993</td>
<td>Natural Language Annotation</td>
<td>Geography, MIT InfoLab</td>
<td>Precompiled knowledge base</td>
<td>NO answer extraction</td>
</tr>
<tr>
<td>Webclopedia</td>
<td>2000</td>
<td>CONTEX parser</td>
<td>Synset based</td>
<td>WordNet</td>
<td>manually built QA patterns, patterns are not robust</td>
</tr>
<tr>
<td>IONAUT, Abney et al.,</td>
<td>2000</td>
<td>own crawler to index</td>
<td>exact answers and Snippets</td>
<td>focus on entities</td>
<td>Snippets support is problem</td>
</tr>
<tr>
<td>Pisa Question Answering System</td>
<td>2001</td>
<td>semantic filters,</td>
<td>WNSENS</td>
<td>retrieved paragraphs</td>
<td>non-factoid based questions missing</td>
</tr>
<tr>
<td>LCC</td>
<td>2001</td>
<td>Syntatic Parser</td>
<td>Knowledge axioms</td>
<td>paragraphs</td>
<td>Monotonic reasoning</td>
</tr>
<tr>
<td>ASKMSR, Brilletal.,</td>
<td>2002</td>
<td>regular expression patterns</td>
<td>Answer Redundancy</td>
<td>Fact-based, short-answer questions</td>
<td>Decision tree</td>
</tr>
<tr>
<td>AnswerBus, Zheng,</td>
<td>2002</td>
<td>Sentences based</td>
<td>Open Domain</td>
<td>Sentence level Web IR</td>
<td>Performance would decrease if exact answers were being evaluated</td>
</tr>
<tr>
<td>AnswerFinder, Qi et al.,</td>
<td>2002</td>
<td>NSIR search engine</td>
<td>Returns ranked exact answer</td>
<td>no context is provided</td>
<td>user still has to read the original document to verify that a given answer is correct</td>
</tr>
<tr>
<td>QAYSO, Abdullah</td>
<td>2011</td>
<td>conceptual hierarchies</td>
<td>Yet Another Great Ontology (YAGO)</td>
<td>WordNet dictionaries and Wikipedia</td>
<td>linguistic triple</td>
</tr>
<tr>
<td>Webquestions berant,</td>
<td>2013</td>
<td>Ontology based</td>
<td>Close domain</td>
<td>wh-word</td>
<td>single source of knowledge</td>
</tr>
<tr>
<td>Universal Network Language, Kanu Goel</td>
<td>2016</td>
<td>IR graph based pattern matching</td>
<td>Close domain</td>
<td>Factoid</td>
<td>non-factoid based questions, Structured DB</td>
</tr>
</tbody>
</table>
Chapter 3

DELAY ASSESSMENT IN QUESTION ANSWERING SYSTEM

This chapter briefly illustrates the delayed assessment in the existing LMS based system used to answer the student’s questions. This system is manual and dependent on the instructors. If the case, either LMS is not working or instructor is not available, then the student’s should face the ultimate delay in response to their questions. This delay should be measured and minimized for the student’s satisfaction and learning curve.

3.1 Question Answering in LMS

In the eLearning system, question answering is more demanding than traditional environments. In the absence of an instructor, students may take their lectures online and after taking their lectures required more question to be addressed. The question answering in eLearning paradigm is deferred on the availability of the teacher. The students have Moderated Discussion Boards (MDBs) for the questioning about the lecture content and other subject related problems. MDBs is manual system LMS based system that currently used to facilitate the student’s question answering as shown in Figure 3.2.

Conceptualize the scenario comprising of LMS as shown in Figure 3.1, in which a students take online course lectures. Suppose the students faces a delay in answering their posted question on MDBs. They may be able to get a response after few hours or perhaps a day, during that time the student remains unable to proceed, confused,
and disappointed. This leads to a decrease in education quality and ultimately student participation in MDBs may suffer. For our investigation, deciding about a suitable examination approach is essential, since it will help us reach specification with respect to our concerned questions.

Moreover, this chapter focuses to the determination of the investigation technique will help us to diagram the explanations and issues relating to our question in the existing system. Delays in answering are complicated and complex, and it would be impossible to select an examination method that is within the limitations of what we can perform. A qualitative examination method offers results that are typically rich and distinct, demonstrating an enormous amount of arrangements. A qualitative investigation strategy has the ability to instruct us on how people feel and what they acknowledge.

### 3.2 Delay Factor in Existing System

The existing question answering system is based on moderated discussion boards and equally visible for the student and teachers. To calculate the delay in responding to student’s questions.
3.2.1 Data Acquisition

The data acquisition is done from the real time MDBs data. There are more than three hundred and thirty four pages data is extracted with each page ten student questions and instructor replies. The students posted their questions at MDBs and that is answered by the instructor as shown in Figure 3.2. Each page contains the student question and teacher replies. The data is preprocessed for the analysis. The anomalies are removed from the collected data. The Algorithm 3.1 presents the step taken for data pre-processing and analysis.

3.2.2 Data Pre-Processing

The data is collected initially in an unstructured format and pages contain Hyper Text Markup Language (HTML) tags. These HTML pages also contain Cascading Style Sheets (CSS) code, which is used for formatting these documents for the aesthetics look and feel of web documents. The data contains the empty lines, symbols, extra characters and irrelevant text. This data is pre-processed to eliminate the log information, CSS, and page structure information as denoted in Algorithm 3.1. Figure 3.3 presents the data analysis model for the delay in the QA System eLearning paradigm.

After the removal of the non-related data, the analysis is started. The R tool version 3.2.5 for Statistical Computing Platform 64-bit is used. Some of the results are also transformed using the Tableau Desktop Software (version 9.3). The results are discussed...
in detail in next section. The aim of this study is to analyze the student motivation towards posting questions and measures the responses delay by the teacher in eLearning paradigm.

### 3.2.3 Data Analysis

Student question analysis in perspective to the question text and delay measurements are extremely useful for the student engagement [88, 89]. The responses are more motivated towards the learning and effectiveness of the online learning [90]. This leads the student to misconceptions in learning concepts [91]. The basic purposes is to enhance the optimum usage of the resources [92]. Learning Management System (LMS) is used in eLearning environments to facilitate the student and institute interactions. The students are allowed an interface name as Moderated Discussion Board (MDB) for the questioning. The instructors have obligatory to answer these questions.

The data is collected about the different 45 lectures. The data acquisition is done from the real time MDBs data. There are more than three-hundred-pages of data is extracted. Each page contains the ten student question and teacher replies. The data is preprocessed for the analysis. The collected data is unstructured and pages are transformed. The data contains the empty lines, symbols, extra characters and irrelevant text. Figure 3 presents the Tableau query for the data clustering used for the analysis.

#### Query 3.1: Tableau Query for the Data Clustering

1. (restrict

2. (aggregate
3. (leftjoin
4. (table [TableauTemp],[data2modified3#csv])
5. (restrict
6. (table [TableauTemp],[data2modified3#csv])
7. ([StudentID])
8. )
9. (([Delay] [StudentID]))
10. )
11. ([StudentID] [StudentID])
12. (sum:Delay:ok
13. (sum [Delay])
14. )
15. (sum:Number of Records:ok
16. (sum 1)
17. )
18. )
19. )
20. ([StudentID] [sum:Delay:ok] [sum:Number of Records:ok]
21. )

3.3 Algorithm Pseudo Code for the Delay Data Analysis

The word cloud is formed based upon the frequency of the words appeared in the student questions. The formation of the word cloud is based on the top number frequencies with which the words appear. Figure 3.4 represents the wordcloud of the extracted text from the question.

Algorithm 3.1: Pseudo Code for the Delay Data Analysis

Input: Comma Separated Values data file (*.CSV)
Output: Clusters of the student question-answer delay

1. Start

2. Start Preprocessing

3. Remove punctuation marks from the text

4. Remove numbers

5. Convert the text to lowercase

6. Remove "stop words"

7. Remove common word endings* (e.g., "ing", "es")

8. Strip whitespace

9. End Preprocessing

10. Stage the data and remove anomalies

11. FOCUS - on just the required stuff

12. Start by removing sparse terms

13. Start Calculating Word Frequency

14. Initialize a two column matrix

15. Count each term and put in a matrix

16. Words appear and the bottom number reflects how many words appear that frequently

17. Considering the threshold greatest frequencies

18. View a table of the terms after removing sparse terms, as above

19. The above matrix was created using a data transformation

   Comment: An alternate view of term frequency:

   Comment: This will identify all terms that appear frequently
   (in this case, 50 or more times)

   Comment: Change "50" to whatever is most appropriate for your data

20. Plot Word Frequencies that appear as the threshold times
21. End Calculating Word Frequency

22. Start Finding Relationships

23. Find Term Correlations

24. If words always appear together, then correlation=1.0

25. Change ”question” & ”analysis” to terms that actually appear in your texts

26. Also, adjust the ‘corlimit’ to any value that is necessary

27. End Finding Relationships

28. Start Word Clouds

Comment: First load the package that makes word clouds in R

29. Perform Hierarchal Clustering by using term similarity and word frequencies

30. End Word Clouds

31. Start K-means Clustering

32. First calculate distance between words

33. Now group them by the number of clusters

34. Draw Dendrogram around the clusters

35. Perform K-means clustering

36. End K-means Clustering

37. End

3.3.1 Trends in Wordcloud

The words with same frequency are shown in same font size and color as shown in the Figure 3.4.

It may also be used to find the trends in question answer about which specific concept question are asked. Given word cloud represent the trends about the most of the queries are found the word assignments means that most of the time students face the difficulties while preparing a solution of their assignments.
3.4 Statistical Analysis of the Collected Data

Table 3.1 represents the summary of the collected data in the form of descriptive statistics. Total extracted data records are 3434 in the dataset. Each record represents the student question with postdate. The date on which the student put the question is called posting date. The teacher reply time and date are also included in the dataset.

A total number of student’s posted questions during the whole semester are '1108’. The enrollment of the student in this course is about '5,000’. Standard deviation is '136.348’ hours from the collected data sample. The maximum delay calculated is 1472 hours for a single student and a minimum delay count is usually 1 hour.

The delay in minutes is negligible because the teacher always takes the time to read and understand student question. Few minutes also required to the teacher for compiling and postin the question reply. The average delay is measured for each student is three and half days, which affects the learning. It is a significant number for the students to wait for answers. Total delay is 93546 hours that is transformed into 3897 days in answering all the questions of the students. This delay is larger than the total days required to complete a semester for any single student.
3.5 Students Cluster versus Delay in Hours

The clustering is performed to group the similar delays as shown in Figure 3.5. The size the bubbles show the size of each cluster and number in it represents the value of delay. Total clusters are formed during analysis of the data are 102. In other words, the total number of the students is combined in the same cluster, whose delay time is same.
The delay time is more in small size cluster, while less in big clusters. The minimum delay is 1 and maximum delay is 714 hours and median is 12.00. In some cluster the delay time to response the student questions are close to hundred hours’ delay. The average delay for clustered data is 33.67 hours. This delay leads the student to the annoying situation and stops the learning process or at least the hindrance in this process.

3.6 Number of Records versus Sum of Delays

The student feels helpless when more than 1-hour delay occurs to reply them.

![Figure 3.6: Number of Records versus Sum of Delays](image)

Figure 3.6 presented the sum of the records along the x-axis. While the sum of the delays is shown along the y-axis. The sum of the delays is the dependent variable, which shows the delay value from 0 to more than 1400 delays.
3.7 Delay Against Distinct Student IDs

Figure 3.7 shows the delay against each distinct Student ID. For the whole semester, the cumulative delay of the student is represented on the bubbles in hours. These delays occur in answering the single student questions during one semester.

3.8 Correlation Representation

Correlation denotes the general class of statistical association comprising dependency, although in collective treatment it most often discusses that two variables at the extent have a linear relationship.

Figure 3.8 represents the correlation among the terms represented by the instructor answers in response to a student question. There 3434 records data are after preprocessing is analyzed. The question terms are plotted along X-axis and answering terms are plotted along the Y-axis. The total terms are added as an appendix in the thesis.
3.9 Conclusion

This delay analysis is strong evidence for administration to show the student teacher question answering gap. Now it is clearly helpful for the administration of the institutions to check the student teacher ratio in both current and future semesters. This research will also be useful to reduce the load on teachers and stress of students. This analysis is motivated the author for research to model automated solution for the student question. The delay factor must be reduced in the novel solution. In the next chapter author propose a framework to address this delay in the responses of the student’s questions.
Chapter 4

MULTIMEDIA AND AGENT BASED QUESTION ANSWERING SYSTEM FOR ELEARNING

The proposed question answering systems rely on document retrieval and information extraction. Retrieved documents which are likely to contain an answer to a student’s question. A question answering system heavily depends on the effectiveness of an information retrieval system. In case, if an information retrieval system fails to find out any relevant documents for a question, further processing steps to extract an answer will inevitably fail as well. This chapter presents the proposed architecture and techniques with respect to their usefulness for question answering.

4.1 Question Answering Architecture

Different Question Answering Systems have some common functions. There are five common steps in general QA architecture for giving the proper answers to the students. These steps are: a) Classification of Questions: Questions are classified into different categories. b) Extracting of Keywords: keywords are extracted from the question statements and extracted documents. Keywords are the special key terms which are used to extract the required sentences that are relevant to the posted question. c) Web Documents Extraction: required documents are retrieved from the different web pages.
d) Answers Extraction: different techniques are made in order to extract the relevant answers from the retrieved documents. e) Answers Ranking: the different answers are extracted from the web documents, then these extracted answers are given the rank number so that most relevant answers can appear at the top in the list of candidate answers [93]. Then the top ranked answer delivered to the students.

4.2 Proposed “Question Answering (QA) System” Architecture

The block diagram for the proposed “Multimedia and Agent-based Question Answering System (MAQAS)” is given in Figure 4.1.

The proposed model in figure 4.2 presents the setup for eLearning system.

There are eight modules in proposed Question Answering System. These eight modules are followings as shown in Figure 4.3:

1. Question Categorization Module (QC Agent)
2. Question Correction Module (QCo Agent)
3. Keywords Extraction Module (KE Agent)
4. Question and Answers Identification Module (QAI Agent)
5. Image-based Answers Module (IA Agent)
6. Answer Extraction & Processing Module (AEP Agent)
7. Feedback and Relevancy Module (FR Agent)
8. Appointment Fixing Module (AF Agent)

The overall functionality of the system is described by integrating different module, which are designed and developed for the proposed solution. The MAQAS is a Smart Framework that manages the experimental setup by integrating the components: corpus, MAS, knowledge repository, NLP tools, image annotator, web document, UML vocabulary, question processing module, answer ranking module, student, and teacher as shown in Figure 4.3. Different types of answers related to QA systems are like factoid, description, definition, ratings. The eLearning environment proposed that combine
a MAS and semantic web technology in order to achieve better search results and task workflows. SWS, MAS and eLearning are very active areas where more work is required to complete the automation of the solution work-flow and a task-centric environment. We will try to build the system by combining these strengths and mechanisms, and attempt to refine the functionalities of multi-agents on semantic web technology in an eLearning environment.

Agent system design and development are mainly based on the use of the DB system. An agent uses the DBs in two ways: 1) a DB is used externally for user-information extraction; 2) a DB is used internally for the storage of information to build user profiles.
Agents are required for the execution of question triggers, stored procedures, and user-defined functionalities. If the intelligence level is very high, if more resources like the operating system, databases, and servers are used, then the budget limits must be increased. Maintenance is also required for the software. In addition to the maintenance of the software agents, the DBs storage also require maintenance and support services. A corpus is used in LMS, as it is more convenient for agent intelligence implementation.

For developing the MAQAS system, different development framework and tools are used. These development language, framework and tools are Python, Python IDLE 3.4, NLTK tool, WordNet and Alchemy API tool. MOQA was designed and developed to work with an LMS. We have three main modules: LMS, MOQA and LMS-MOQA interface module (Figure 4.4) LMS can be any LMS providing online learning services to learners. MOQA is the multi-agent system. It has several agents to perform the adaptive services required by LMS.

The student enters the question related to their subject and what would they the problem they are facing. The questions are posted by using the interface provided by the question answering framework. The question by the students are initially categorized into either text based or image based question. The different modules are available for handling the both type of questions. The text type questions pass out the spell checker and statement
Figure 4.3: Multimedia and Agent based Question Answering System (MAQAS)
of the questions is confirmed by some suggestion and options. In case of question’s statement contains some errors then corrected. The question is tokenized and parsed for part of speech tagging. The keywords are extracted from the parsed statement then among these keywords headwords are identified. These headwords are main the focus of the student question and handled carefully for relevant answer extraction. Now the keywords and headword are used for the document retrieval form the web sources. The question words are used for the identification of question and answers type.

Document retrieval is carried out using three well-known search engines: Google, Wikipedia and Webopedia. The top ten links are retrieved from Google and Wikipedia for the document retrieval. The text retrieved from the relevant links and saved into text documents. These text documents are used for the answer extraction and processing. Various corpuses are used for the extraction of relevant answers from the text documents. The questions are more accurate answered by the help of their answer type which is already identified.

The extracted answers are further processed for the ranking according to the relevance with a question. The top most ranked answer is delivered to student’s and feedback is collected from the student. If the student feedback is satisfactory then the process will store the question answer into repository for future usage. On the other hand, if the students feedback is unsatisfactory then appointment module will active and collaborate with teacher and student for suitability of appointment slot. The teacher available slot checked and students location is identified by collecting coordinates using global positioning system. Student appointment is fixed with respect to teacher available slot and students consent for the multimedia content delivery. The detail of the components are discussed in next section.

4.3 Development Tools

For developing the MAQAS system, different development framework and tools are used. These development language, framework and tools are Python, Python Integrated Development and Learning Environment 3.4, Nature Language Toolkit, WordNet and Alchemy API tool.
4.3.1 Python

Python is a language in which different mature applications are developed. This language is easy to use and for the quick development. Many developers and researchers are using the Python language throughout the world for their scientific research. So, Python language is used in order to develop the MAQAS System integrated with JAVA.

4.3.2 WordNet

WordNet dictionary is a large volume of the lexical database that contains the different type of word. This word contains the noun phrases, verb phrases, adverbs, conjunctions, disjunctions, adjectives are grouped into the cognitive set as synsets. WordNet dictionary is used for the spelling correction of the student’s posted questions. The students questions are reformulated if there exist any spelling error.

4.3.3 Natural Language Toolkit (NLTK)

There are different Natural Language Processing tools for understanding the human languages. Nature Language Processing tools is NLTK used for question statement processing. Since Natural Language Toolkit is the efficient and excellent tool for language processing. So, NLTK tools are used in MAQAS system for tokenizing the question into words, parsing of the question, and the part of speech tagging for extracting answers from the documents.

4.3.4 Alchemy Keyword Extraction API

Alchemy API is also NLP tool that is developed by a famous company Alchemy API. By using Alchemy API, keywords are extracted from the questions and the retrieved documents in order to extract the most relevant answers to the posted question. Alchemy API is also used for extracting the named entities (NE) from the documents.

4.4 Multi-agent Object Oriented Question Answering (MOQA)

Multi-agent Object Oriented Question Answering (MOQA) is developed and used for the MAQAS system. MOQA connector provides the connections between the Learning Management System interface and Learning Management Agents System as shown in Figure 4.4.

MOQA was designed to work with an LMS. We have three main modules: LMS ,
Figure 4.4: Architecture of Multi-agent Object Oriented Question Answering (MOQA)

MOQA and LMS-MOQA interface module as shown in Figure 4.3. LMS can be any LMS is providing online learning services to learners. MOQA is the multi-agent system. It has several agents to perform the adaptive services required by LMS. LMS-MOQA interface is the communication platform of these two separate modules. A socket-based communication protocol is used to achieve communication. MOQA can communicate with any LMS, if the LMS sends the data packets in the same format defined in the protocol as shown in Table 4.1.

The system has LMS Agent, QuestionAnswering (QA) Agent with seven learning agents: QuestionCategorization (QC) Agent, QuestionCorrection (QCo) Agent, KeywordExtraction (KE) Agent, QuestionAnswerIdentifier (QAI) Agent, AnswerExtraction and Processing (AEP) Agent, Feedback and Relevvancy (FR) Agent and AppointmentFixing (AF) Agent. The descriptions and roles of each agent are as follows:

1. LMSAgent is the communication party with the LMS. It behaves as the MOQA server. It receives MOQA request from the LMS part and provides request to the
Table 4.1: Responsibilities and Activities of LMS Agent

<table>
<thead>
<tr>
<th>Role: LMSAgent (LMSA)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong> LMS Agent is the communication party with the LMS. It behaves as the MOQA server. It receives MOQA request from the LMS part and provides request to the QuestionAnswering Agent. It receives results from the QuestionAnswering Agent and provides MOQA Respond message to the LMS.</td>
</tr>
<tr>
<td><strong>Protocols:</strong> SetLMS, SetMOQA,</td>
</tr>
<tr>
<td><strong>Activities:</strong> ActiveLMSfunction, ActiveMOQAfunctions,</td>
</tr>
<tr>
<td><strong>Permissions:</strong> reads acquaintances</td>
</tr>
<tr>
<td><strong>Responsibilities and liveness:</strong> LMSA= (SetLMS, SetMOQA, ActiveLMSfunction, ActiveMOQAfunctions).</td>
</tr>
<tr>
<td><strong>Safety:</strong> A successful connection with the database and LMS is established</td>
</tr>
</tbody>
</table>

Figure 4.5: Gaia Design and Analysis Model for the Agents [2094]

QuestionAnswering Agent. It receives results from the QuestionAnswering Agent and provides MOQA Respond message to the LMS as shown in Table 4.2.

2. QuestionAnswering Agent is the central agent which is responsible for management of the other agents and communication with LMS Agent as shown in Table 4.3.

3. QuestionCategorization Agent is responsible for finding the most appropriate content for the student using the questions corpus and dictionaries. It also communicates with QuestionCorrection Agent.

4. QuestionCorrection Agent is trained for the auto correction of the statements of the question posted by the student. It further communicates with KeywordExtraction Agent.
5. KewordExtraction Agent used the different techniques and found the keywords from the question statements. After extracting keywords classification and matching, it sorts and filters for the headwords identification. These keywords and headwords are used in search engines. It further communicates with QuestionAnswering Agent.

6. QuestionAnsweringIdentification Agent used the question word for the identification of question and answer types. It further used to analyze the type of the answers required by the questions. This agent communicates with AnswerProcessing Agent.

7. AnswerExtractionsProcessing Agent receives result of the keywords and headwords extracted by the KE Agent. It also receive result of the question type and answers types identified by the QAI Agent. It used the three type of search engines to retrieve the passages containing the answer from the web sources. It also communicates with Feedbackandrelevancy Agent.

8. FeedbackandRelevancy Agent collects the feedback from the student and relevancy score of the answer contents. It further communicates with Appointment Agent.

9. Appointment Fixing Agent receives the result from the FeedbackandRelevancy Agent. It also collects location coordinates from the GPS to pinpoint the student location. To keep in view teacher free slots, the appointment agent fixes the appointment of the student and teacher for synchronous communication.

The agents in MOQA were developed using JADE agents as shown in Figure 4.6. MOQA used the GAIA model for analysis and design as depicted in Figure 4.5 using Java Agent Development Environment (JADE) [19] is a software development framework aimed to develop multi-agent systems and applications conforming to Foundation for Intelligent Physical Agents (FIPA) standards for intelligent agents. For more information about FIPA \[^{14}\], JADE is one of the most commonly used agent middleware and it has well-structured agent management mechanisms providing a runtime environment, a library of classes that programmers can use, and a suite of graphical tools that allow administrating and monitoring the activity of running agents.

[^14]: [www.fipa.org](http://www.fipa.org)
Chapter 4. Multimedia and Agent Based Question Answering System for eLearning

Figure 4.6: Generalized Architecture of the System

Figure 4.7: Question Posting Interface for Students
4.5 Student Question Posting Interface

The student enters the question related to their subject and what would be the problem they are facing. These questions are posted by using the interface provided by the question answering system. The module as shown in Figure 4.8 depicts the work flow of the question categorization. The interface screen shot is shown in Figure 4.7.

![Diagram of Student Question Posting Interface](image.png)

**Figure 4.8**: Question Categorization Module

4.6 Student Question Categorization

Let us consider a comprehensive level scenario in which a student takes a lecture presenting an introduction to software engineering. The student can enter a question in different ways. The student is allowed to enter the English language questions at the initial level. In addition to that student may enter the Unified Modeling Language(UML) images, a diagram as a question to get information about the said image and diagram.
Chapter 4. Multimedia and Agent Based Question Answering System for eLearning

Table 4.2: Responsibilities and Activities of QA Agent

<table>
<thead>
<tr>
<th>Role: QuestionAnsweringAgent (QAA)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong> QuestionAnswering Agent is the central agent which is responsible for management of the other agents and communication with LMS Agent.</td>
</tr>
<tr>
<td><strong>Protocols:</strong> SetLMS, SetMOQA, SetQAA,</td>
</tr>
<tr>
<td><strong>Activities:</strong> ActiveLMSfunction, ActiveMOQAfunctions, ActiveQAFunction</td>
</tr>
<tr>
<td><strong>Permissions:</strong> reads acquaintances</td>
</tr>
<tr>
<td><strong>Responsibilities and liveliness:</strong> LMSA = (SetLMS, SetMOQA, ActiveLMSfunction, ActiveMOQAfunctions, ActiveQAFunction)</td>
</tr>
<tr>
<td><strong>Safety:</strong> A successful connection with the database and LMS is established and sub-agent active.</td>
</tr>
</tbody>
</table>

Table 4.3: Responsibilities and Activities of QC Agent

<table>
<thead>
<tr>
<th>Role: QuestionCategorization Agent (QCA)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong> QuestionCategorization Agent is responsible for finding the most appropriate content for the student using the questions corpus and dictionaries. It also communicates with QuestionCorrection Agent.</td>
</tr>
<tr>
<td><strong>Protocols:</strong> SetLMS, SetMOQA, SetQCA</td>
</tr>
<tr>
<td><strong>Activities:</strong> ActiveLMSfunction, ActiveMOQAfunctions, SetQCAFunction</td>
</tr>
<tr>
<td><strong>Permissions:</strong> reads acquaintances</td>
</tr>
<tr>
<td><strong>Responsibilities and liveliness:</strong> LMSA = (SetLMS, SetMOQA, ActiveLMSfunction, ActiveMOQAfunctions, SetQCAFunction)</td>
</tr>
<tr>
<td><strong>Safety:</strong> A successful connection with the database and LMS is established</td>
</tr>
</tbody>
</table>

Students are also provided the facility to enter both text questions and images concurrently. MAQAS will answer the student questions. The student enters the question and the question answering system starts to first all to categorize the student questions as shown in Figure 4.8.

In this module, questions are categorized into different categories. When a student enters a question in an MAQAS system, then the question is sent to this module. In which posted question is categorize that either question is the type of “Text” or “image”. After categorization process, posted question is sent to their respective agent. If the posted question is the type of “Text Type”, then it is forwarded to the “QCo Agent” that is “Question Correction Module”.

On the other-hand, if the posted question is “Image” instead of text, then it is sent to the “IA Agent” that is “Image-based Answers Agent”. Student question is categorized into two categories as shown in Figure 4.9.

1. Text Questions
2. Image Questions
4.6.1 Text Questions

As the system identifies the category of the questions, then agents are activated with each question type. For example, if the student enters the text question, then QCo Agent handled this question for correction.

4.6.2 Image Questions

In case, if a student enters the image instead of question text, then IA Agent takes the responsibility to tackle this question and process it further. If the student chooses the topic ‘UML use case diagrams,’ and then wants information about the use case relation for ATM Transactions by entering the diagram of ATM Use Case. If the entered question is hybrid and contains both text and image, then IA Agent handle it for further processing. A student may ask a question about use cases, e.g., “What is the main purpose of the use case?” Alternatively, the student may enter an image or diagram and request a description of it. The student may also enter a question along with a diagram. Finally, the student may enter an image-based question in which the image displays a text rather than a diagram.
4.7 Question Correction Module (QCo Agent)

The student entered a question and is categorized as text. Then the QCo Agent tackles the entered question by checking the syntax of the question. Agent collaboration in question answering system is shown in Figure 4.10.

To check the question is syntactically correct or incorrect, WordNet dictionary is used. The statement of the question entered by the student are word by word matched with the WordNet.

If the words in the question that are misspelled will be highlighted and correction is suggested by the pop-up message. This correction leads toward the question reformulation partially are complete. The QCo Agent handled the questions to correct for further processing. Figure 4.11 represent the question handled by the QCo Agent for the reformulation using the dictionary. In this module, posted questions are corrected before sending to other modules for further processing. Responsibilities and activities QuestionCorrection Agent is shown in Table 4.4.

As long as a posted question is received then the QCo Agent checks the grammatical syntax errors in the question by using WordNet dictionary. If any grammatical error is found in the question, then it will be corrected. For example, the student posted a question about the computer history and due to misspelled the question becomes the
Figure 4.11: Question Statement Correction Module

Table 4.4: Responsibilities and Activities Question Correction Agent

<table>
<thead>
<tr>
<th>Role:</th>
<th>Question Correction Agent (QCA)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>Question Correction Agent is trained for the auto correction of the statements of the question posted by the student. It further communicates with Keyword Extraction Agent.</td>
</tr>
<tr>
<td><strong>Protocols:</strong></td>
<td>SetLMS, SetMOQA, SetQCoA</td>
</tr>
<tr>
<td><strong>Activities:</strong></td>
<td>ActiveLMSfunction, ActiveMOQAfunctions, SetQCAFunction</td>
</tr>
<tr>
<td><strong>Permissions:</strong></td>
<td>reads acquaintances</td>
</tr>
<tr>
<td><strong>Responsibilities and liveness:</strong></td>
<td>LMSA = (SetLMS, SetMOQA, ActiveLMSfunction, ActiveMOQAfunctions, SetQCAFunction)</td>
</tr>
<tr>
<td><strong>Safety:</strong></td>
<td>A successful connection with the database and LMS is established</td>
</tr>
</tbody>
</table>
“who was the inventor and further of the computer”. The question is wrong and needs to correct. Now QCo Agent checks the spelling errors in the question by using WordNet dictionary, if any spelling error is found then a Pop-Up message will appear. The agent one with the help of the WordNet dictionary corrects the student’s questions. In this pop-up message, different suggestions will be given for correcting the spelling error or given the proper synonyms words. When a student corrects the spelling error from the question, then this question reformulates according to their correct grammatical syntax and spellings. So, now this question is called the “Corrected Question” that is ready for further processing in KE Agent.

The frequency of errors in spelling in the human typed text is estimated from 1-2% for retyping already printed text carefully to 10-15 % for web based questions. Spelling fixation is a necessary part of modern world’s writing. As writing is part of texting, composing emails, writing documents and web surfing. Current spelling correctors are not perfect and they are pervasive pretty much in any software that has keyboard input.

4.7.1 Spelling Correction Types

There are two types of spelling correction types: Non-Word Correction and Real word correction.

4.7.1.1 NonWord Correction

Non-word correction is a kind of correction in which a word which does not belong to any word of the dictionary. however, we can see its opposite Real word correction is a type in which corrections are done in correct words. Correction in Non-word is easy, as its detection is also easy, if a word is not found in dictionary then it must be corrected by substitution real world belongs to dictionary.

4.7.1.2 RealWord Correction

Correction in realword is not easy, although its detection is easy. Realword is found in the dictionary then have to consider other factors like sentence flow. We consider here word “buble” as it has its considerations. The edit distance with inversion of row and columns is named as Damerau-Levenshtein edit distance, which is distance between two words is the minimum number of operations (consisting of insertions, deletions or substitutions of a single character, or transposition of two adjacent characters as shown

Table 4.5: Candidate Corrections for the Misspelled Word "buble"

<table>
<thead>
<tr>
<th>Error Word</th>
<th>Corrected Word</th>
<th>Correction Letter</th>
<th>Letter</th>
<th>Letter#</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>buble</td>
<td>double</td>
<td>do</td>
<td>b</td>
<td>0</td>
<td>Substitution</td>
</tr>
<tr>
<td>buble</td>
<td>bubble</td>
<td>b</td>
<td>-</td>
<td>2</td>
<td>Deletion</td>
</tr>
<tr>
<td>buble</td>
<td>trouble</td>
<td>tro</td>
<td>bu</td>
<td>0, 1</td>
<td>Substitution</td>
</tr>
<tr>
<td>buble</td>
<td>table</td>
<td>ta</td>
<td>bu</td>
<td>0, 1</td>
<td>Substitution</td>
</tr>
<tr>
<td>buble</td>
<td>noble</td>
<td>no</td>
<td>bu</td>
<td>0, 1</td>
<td>Substitution</td>
</tr>
<tr>
<td>buble</td>
<td>hubble</td>
<td>h, b</td>
<td>b</td>
<td>0, 2</td>
<td>Substitution, Deletion</td>
</tr>
<tr>
<td>buble</td>
<td>able</td>
<td>a</td>
<td>bu</td>
<td>0</td>
<td>Substitution</td>
</tr>
</tbody>
</table>

in Table 4.8. It is applicable for the single transformation and reformulation of the questions. The correct candidate word are given below and then transformed into the correct words. Candidate corrections for the misspelled word "buble" and "futher" as shown in Tables 4.5 and 4.6.

4.7.2 NOISY CHANNEL Spelling Correction

Noisy channel algorithm [63] is used for the word correction is given as Algorithm 4.1.

Algorithm 4.1: NOISY CHANNEL Spelling Correction

1. function NOISY CHANNEL SPELLING(word x, dict D, lm, editprob)
2. returns correction if x $\notin$ D
3. candidates, editsAll strings at edit distance 1 from x that are $\in$ D, and their edit for each c, e in candidates, edits
4. channeleditprob(e)
5. priorlm(x)
6. score[c] = log channel + log prior
7. return argmaxc
8. score[c]

The candidate corrections for the words misspelling "futher" and "buble" would be transformed.

The Corpus of Contemporary English of America (COCA) having words 404,253,213 for uni-gram model.
Chapter 4. *Multimedia and Agent Based Question Answering System for eLearning* 68

### Table 4.6: Candidate Corrections for the Misspelled word "futher"

<table>
<thead>
<tr>
<th>Error Word</th>
<th>Corrected Word</th>
<th>Correction Letter</th>
<th>Letter</th>
<th>Letter #</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>futher</td>
<td>father</td>
<td>a</td>
<td>u</td>
<td>2</td>
<td>Substitution</td>
</tr>
<tr>
<td>futher</td>
<td>further</td>
<td>r</td>
<td>-</td>
<td>2</td>
<td>Deletion</td>
</tr>
<tr>
<td>futher</td>
<td>rather</td>
<td>ra</td>
<td>fu</td>
<td>0,1</td>
<td>Substitution</td>
</tr>
<tr>
<td>futher</td>
<td>other</td>
<td>o</td>
<td>fu</td>
<td>0,1</td>
<td>Substitution</td>
</tr>
<tr>
<td>futher</td>
<td>gather</td>
<td>ga</td>
<td>fu</td>
<td>0,1</td>
<td>Substitution</td>
</tr>
<tr>
<td>futher</td>
<td>future</td>
<td>u re</td>
<td>h er</td>
<td>3 4,5</td>
<td>Substitution Transposition</td>
</tr>
</tbody>
</table>

### Table 4.7: Word with Reasonable Weight Estimation

<table>
<thead>
<tr>
<th>Word</th>
<th>Part of speech</th>
<th>Frequency</th>
<th>Dispersion</th>
</tr>
</thead>
<tbody>
<tr>
<td>father</td>
<td>n</td>
<td>145051</td>
<td>0.93</td>
</tr>
<tr>
<td>rather</td>
<td>r</td>
<td>44934</td>
<td>0.96</td>
</tr>
<tr>
<td>further</td>
<td>r</td>
<td>10349</td>
<td>0.86</td>
</tr>
<tr>
<td>other</td>
<td>j</td>
<td>547799</td>
<td>0.97</td>
</tr>
<tr>
<td>gather</td>
<td>v</td>
<td>27659</td>
<td>0.97</td>
</tr>
<tr>
<td>future</td>
<td>n</td>
<td>58020</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Pretty reasonable estimate of P(xjw) just by looking at the local context: the identity of the correct letter itself, the misspelling, and the surrounding letters. Word with reasonable weight estimation is represented in the Table 4.7.

The model used for the estimation, for example, p(xjw) as P(futher \(\mid\) father) just using the number of times that the letter a was substituted for the letter u in some large corpus of errors.

Thus for example a substitution matrix will be a square matrix of size 26 26 (or more generally \(j \times j\), for an alphabet A) that represents the number of times one letter was incorrectly used instead of another.

\[
\begin{align*}
\text{del}[x; y]: & \quad \text{count}(xy \text{ typed as } x) \\
\text{ins}[x; y]: & \quad \text{count}(x \text{ typed as } xy) \\
\text{sub}[x; y]: & \quad \text{count}(x \text{ typed as } y) \\
\text{trans}[x; y]: & \quad \text{count}(xy \text{ typed as } yx)
\end{align*}
\]
Table 4.8: Probabilities for the Words Insertion, Deletions, Transposition and Substitution

<table>
<thead>
<tr>
<th>Words</th>
<th>POS</th>
<th>Count</th>
<th>P(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>father</td>
<td>n</td>
<td>145051</td>
<td>3.59 e -4</td>
</tr>
<tr>
<td>rather</td>
<td>r</td>
<td>44934</td>
<td>1.11 e -4</td>
</tr>
<tr>
<td>further</td>
<td>r</td>
<td>10349</td>
<td>2.56 e -5</td>
</tr>
<tr>
<td>other</td>
<td>j</td>
<td>547799</td>
<td>1.36 e -3</td>
</tr>
<tr>
<td>gather</td>
<td>v</td>
<td>27659</td>
<td>6.84 e -5</td>
</tr>
<tr>
<td>future</td>
<td>n</td>
<td>58020</td>
<td>1.44 e -4</td>
</tr>
</tbody>
</table>

Table 4.9: Candidate Corrections for the Misspelling Across and the Transformations

<table>
<thead>
<tr>
<th>Error</th>
<th>Correction Word</th>
<th>t</th>
<th>a</th>
<th>ac</th>
<th>ac</th>
<th>c</th>
<th>r</th>
<th>e</th>
<th>s</th>
<th>Insertion</th>
<th>Deletion</th>
<th>Transposition</th>
<th>Substitution</th>
<th>Insertion</th>
</tr>
</thead>
<tbody>
<tr>
<td>across</td>
<td>actress</td>
<td>t</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Insertion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>across</td>
<td>cress</td>
<td>-</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Insertion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>across</td>
<td>caress</td>
<td>ca</td>
<td>ac</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Transposition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>across</td>
<td>access</td>
<td>c</td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Substitution</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>across</td>
<td>across</td>
<td>o</td>
<td>e</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Substitution</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>across</td>
<td>acres</td>
<td>-</td>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Insertion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>across</td>
<td>acres</td>
<td>-</td>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Insertion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that we have the conditioned insertion and deletion probabilities are on the previous character. Now instead of choosing to condition on the following character, we get the confusion matrices.

The surrounding words make it clear that father and not further was the intended word. For this reason, it is important to use larger language models than uni-grams.

Evaluating spell correction algorithms is done by holding out a training, development and test set from lists of errors like those on the sites mentioned above.

The prior probability of each correction P(w) is the language model probability of the word w in context, which can be computed using any language model, from unigram to trigram or 4-gram. The Corpus of Contemporary American English to compute the bigram probabilities for the words actress and cross are given in Tables 4.9 and 4.10.

\[ P(\text{actress/versatile}) = 0.000021 \]
\[ P(\text{whose/actress}) = 0.0010 \]

“Steller and versatile actress whose combination of sass and glamorous.”
Table 4.10: Ranking of Each Candidate Correction using Language Model [63]

<table>
<thead>
<tr>
<th>Candidate Correction</th>
<th>Correct Letter</th>
<th>Error Letter</th>
<th>x/w</th>
<th>P(w)</th>
<th>P(x/w)P(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>actress</td>
<td>t</td>
<td>-</td>
<td>c/ct</td>
<td>0.0000231</td>
<td>2.70000000000000</td>
</tr>
<tr>
<td>cress</td>
<td>-</td>
<td>ac</td>
<td>a/#</td>
<td>0.00000144</td>
<td>7.80000000000000</td>
</tr>
<tr>
<td>caress</td>
<td>ca</td>
<td>ac</td>
<td>ac/ca</td>
<td>0.00000164</td>
<td>2.80000000000000</td>
</tr>
<tr>
<td>access</td>
<td>c</td>
<td>r</td>
<td>r/c</td>
<td>0.0000916</td>
<td>1.90000000000000</td>
</tr>
<tr>
<td>across</td>
<td>o</td>
<td>e</td>
<td>e/o</td>
<td>0.000299</td>
<td>2.80000000000000</td>
</tr>
<tr>
<td>acres</td>
<td>-</td>
<td>s</td>
<td>es/e</td>
<td>0.0000318</td>
<td>100000000000000</td>
</tr>
<tr>
<td>acres</td>
<td>-</td>
<td>s</td>
<td>ss/s</td>
<td>0.0000318</td>
<td>100000000000000</td>
</tr>
</tbody>
</table>

\[
P(\text{across/versatile}) = 0.0010\]

\[
P(\text{whose/across}) = 0.000006\]

Evaluating by the spell correction algorithms then the bigram noisy channel chooses the actress word is more suitable for the corrections.

### 4.7.3 Keywords Extraction Module

In this module, Headwords and Keywords are extracted from the question. For this purpose: 1). First of all posted question is tokenized in words by using NLTK tool. 2). In the second step, punctuation marks i.e. colon, semicolon, full stop, question mark and comma etc are removed from the question. 3). The question is parsed. 4). In fourth step, keywords are extracted from the questions. 5). Fifth, headwords are also extracted from the question. When these processes are completed, then posted question are sent to the QAI Agent, function of the KE Agent is shown in Table 4.11.

The reformulated question is passed to KE Agent which tokenized the question statement. KE Agent after tokenized the question statement remove the punctuation marks and spaces. The chunked of questions are further used to determine the pattern and correlation among the words of the question as shown Figure 4.12.

It further leads to help in the question classification. For example:

**Question 1:** What do you mean by memory devices used in computer?

**Question 2:** What is the purpose of using RAM in computer system?

The questions are tokenized and parsed by the NLTK, then tagged using POS by the alchemy NE and Stanford NER. The tagged and passed question are represented by the Figure 4.13:
4.7.4 Sentence Tokenization Process

In sentence tokenization process, sentences in question statement and the retrieved documents are tokenized into words. Since we need to match each word with the question keywords for finding the required answers. So, for this purpose, NLTK is used for tokenizing the sentences in retrieved documents. Following example shows that how a sentence has been tokenized into words.

**Example:**

Computer software, or simply software, is that part of a computer system that consists of encoded information or computer instructions, in contrast to the physical hardware from which the system is built.

['Computer', 'software', ',', 'or', 'simply', 'software', ',', 'is', 'that', 'part', 'of', 'a', 'computer', 'system', 'that', ',', 'consists', 'of', 'encoded', 'information', 'or', 'computer', 'instructions', ',', 'in', 'contrast', 'to', 'the', 'physical', 'hardware', 'from', 'which', 'the', 'system', 'is', 'built', '.']
Table 4.11: Responsibilities and Activities of KeywordExtraction Agent

<table>
<thead>
<tr>
<th>Role: KeywordExtractionAgent (KEA)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong> KeywordExtraction Agent used the different techniques and found the keywords from the question statements. After extracting keywords classification and matching, it sorts and filters for the headwords identification. These keywords and headwords are used in search engines. It further communicates with QuestionAnswering Agent.</td>
</tr>
<tr>
<td><strong>Protocols:</strong> SetLMS, SetMOQA, SetKEA</td>
</tr>
<tr>
<td><strong>Activities:</strong> ActiveLMSfunction, ActiveMOQAfunctions, ActiveKEAfunctions</td>
</tr>
<tr>
<td><strong>Permissions:</strong> reads acquaintances</td>
</tr>
<tr>
<td><strong>Responsibilities and liveliness:</strong> LMSA= (SetLMS, SetMOQA, ActiveLMSfunction, ActiveMOQAfunctions, ActiveKEAfunctions)</td>
</tr>
<tr>
<td><strong>Safety:</strong> A successful connection with the database and LMS is established</td>
</tr>
</tbody>
</table>

4.7.5 Removal of Punctuation Marks

In this phase, different punctuation marks which are a colon, semicolon, commas and full stops are removed from the retrieved documents. When these punctuation marks are removed, then these documents are now sent to the next processes and modules for further processing in order to extract answers.

4.7.6 Part of Speech (POS) Tagging

In order to extract the answers from the documents, we give the part of speech tagging to each tokenized words of documents. So, for this purpose, we also used the NLTK tool for tagging the part of speech to each word in the documents. Following example shows that how tagging the part of speech to each word in documents. Part of speech annotation is shown in Figure 4.14.

Figure 4.14: Question Part of Speech Annotation

4.7.7 Parsing

Sentences parsed the for the understanding the grammatical relationship between the words of the particular sentence. So that we can extract the most relevant answers from the retrieved documents that having the same relations to the relation of posted
question. So, for finding the correlation between the words of the particular sentence, we also used the NLTK tool in our MAQAS system. Following example shows the parsing of a specific sentence in the document as shown in Figure 4.15.

\[(S
    A/DT
computer/NN
program/NN
is/VBZ
a/DT
collection/NN
of/IN
instructions/NNS
that/WDT
performs/VBZ
a/DT
specific/JJ
task/NN
when/WRB
executed/VBN
by/IN
a/DT
computer/NN
./.)\]

Figure 4.15: Sentence Parsing

4.7.8 Named Entities (NE)

In order to find the named entities from the retrieved documents, we use a new NLP tool that is Alchemy API named entity tool. Since named entity required for finding the answers of factoid type of questions. So, Alchemy API used for named entity tool extractor in MAQAS system. Alchemy API finds the different types in named entity in which some are a date, time, organization names, person names and quantities.

4.8 Keywords Extraction Module (KE Agent)

Since keywords play an important role in QA systems in order to find the related answers from textual documents. So, in order to extract the keyword from question statement.

4.8.1 Keywords Extraction

Since keywords play an important role in QA systems in order to find the related answers from textual documents. So, in order to extract the keyword from textual data, the characteristic keyword extraction algorithms adopt the following procedures:

- Exclude any punctuation marks and stop words from the given textual data
• Extract all the possible words and phrases that can be the possible keywords. So these extracted words are the candidate keywords.

• To calculate the properties for each candidate keyword which identifies that it may be a characteristic keyword. The different properties for selection of typical keywords are the frequency of occurrences for a specific term, position in the document, the length of phrasing words, similarity to other candidate keywords.

Figure 4.17 describes that how characteristic keywords are extracted from the basis of their properties commonly by a number of keywords extraction tools. In the MAQAS system, keywords are extracted by using AlchemyAPI Keywords Extractor tool. Now keywords for each extracted text document are extracted and sent to the next component for further processing.

4.8.2 Headword Extraction

Headword is the focusing word or phrase in a question in which user want to seek information about the specific object. This specific object is the focusing word which is used as a headword in this proposed system. So the first noun and the first verb after the WH or question word in the question is the headword word. Headword plays the main role in extracting the answers for the questions of factoid type. Thus in this component headword of the question is extracted. the headword extracted form the question statement is shown in Figure 4.16. Since headword and keywords of question are used in different algorithms in which many patterns are defined in order to find the type of question and identify the type of Answer. When these processes are completed, then posted question are sent to the QAI Agent.

4.9 Question and Answer Type Identification

In this module, identification of the answers type against the posted question is carried out. For this purpose, QAI Agent collects the keywords and headwords of the posted question. And now checks the question words in the question from the given question corpus. Then now QAI Agent classify the question into two main classes by checking in the question that either question word is the type of either “WH-word”or “Non-Wh word”. the function of the QAI Agent is shown in Table 4.12.
**Wh-Words** are those questions that contain the alphabets “W” and “H” in question statement like what, where, when are shown in Table 4.13. NonWh Words are those question words used to ask a question by using simple questions, like describe, list, identify.

<table>
<thead>
<tr>
<th>Table 4.12: Responsibilities and Activities of QuestionAnsweringIdentification Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Role:</strong> QuestionAnsweringIdentification Agent (QAIA)</td>
</tr>
<tr>
<td><strong>Description:</strong> QuestionAnsweringIdentification Agent used the question word for the identification of question types. It further used to analyze the type of the answers required by the questions. This agent communicates with AnswerExtractionProcessing Agent.</td>
</tr>
<tr>
<td><strong>Protocols:</strong> SetLMS, SetMOQA, SetQAIA</td>
</tr>
<tr>
<td><strong>Activities:</strong> ActiveLMSfunction, ActiveMOQAfunctions, ActiveQAIAfunctions</td>
</tr>
<tr>
<td><strong>Permissions:</strong> reads acquaintances</td>
</tr>
<tr>
<td><strong>Responsibilities and liveliness:</strong> LMSA = (SetLMS, SetMOQA, SetQAIA, ActiveLMSfunction, ActiveMOQAfunctions, ActiveQAIAfunctions)</td>
</tr>
<tr>
<td><strong>Safety:</strong> A successful connection with the database and LMS is established</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4.13: Categorization of WH-Word Type Question</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Types of Questions</strong></td>
</tr>
<tr>
<td>Question Words</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

These main classes are “WH type” and “Non-WH type”. If the question is the “WH type”, then the question is classified into three different categories, which are:

1) Definition Type Questions
2) Description Type Questions
3) Factoid Type Questions.

On the other side, if the question is the “Non-WH type” then the question is also classified into three different categories, which are: 1) Definition Type Questions: 2) Description Type Questions 3) MCQ’s Type Questions.

4.9.1 Question Classification Algorithms

For the identification of the answer types algorithms are developed. These algorithms are used for the different type of questions. These algorithms are used to identify the question words which may be the Whtype or NonWH type.

These question words are helpful in finding the correct answer for the student questions. The Algorithm 4.1 is used for the identification of the answers types for the question containing the what types of question. If what is the question word that is used for the factoid type, description and definition type answers, the classification for the what types questions on the basis of the different word in the question used for the identification of exact answer type.

4.9.2 Question Answer Types

Question type is defined as a certain semantic category of questions characterized by some common properties. The major question types are factoids, list, definition, hypothetical, causal, relationship, procedural, and confirmation questions.

Factoid type question is a that, which usually starts with a Wh-interrogated word (What, When, Where, Who). Factoid question requires an answer as a fact expressed in the text body.

List type question is a question, which requires as an answer a list of entities or facts. A list question usually starts as: List/Name [me] [all/at least NUMBER/some].

Definition type question is a question, which requires finding the definition of the term in the question and normally starts with “What is”.

Descriptive type question which asks for definitional information or for the description of an event, and the opinion question whose focus is the opinion about an entity or an event.

Hypothetical type question is a question, which requires information about a hypothetical event and has the form of What would happen if.

Causal type question is a question, which requires an explanation of an event or artifact, like Why.

Relationship type question asks about a relationship between two entities.

Procedural type question is a question, which requires an answer as a list of instructions for accomplishing the task mentioned in the question.

Confirmation type question is a question, which requires a Yes or No as an answer to an event expressed in the question.

Information source is defined as a collection of information objects (documents, video, audio, text, files or databases) available to the question answering system for extracting answers. After the answer type is identified then the particular algorithm is used for the extraction of the answer. Algorithm 4.1 is used for the identification of answer types for ‘What-Type of questions.

Algorithm 4.1: Identification of Answer Types for ‘What-Type of Question’
**Input:** Question ‘Q’

**Output:** Answer types for the question

1. if question Q contain ‘wh-type’ question word “What”, then

2. {

3. if (question q = = “What + aux verb + [ Keyword / Noun / (JJ+ Noun) ]”), then

4. Answer type is Definition Type;

5. Go to Answer Extraction Algorithm for Definition Type;

6. 

7. // ‘ABC’ means anything in the question q

8. else if (question q! = “What + aux + (Determiner Terms Corpus Text File / “the” word ) + keyword + ABC ”)

9. then

10. Answer type is Factoid Type;

11. Go to Answer Extraction Algorithm for Factoid Type;

12. else if (question q = = “What + (The Reasoning Terms Corpus Text File / caused / cause / reason ) + ABC”) Then

13. Answer type is Description Type;

14. Go to Answer Extraction Algorithm for Description Type;

15. else if (question q = = “What / What’s + head words / ( Keyword / ABC) / head words;

16. then

17. // Head word is the first [Noun / Noun Phrase] after WH-word

18. Answer type is Factoid Type;

19. Go to Answer Extraction Algorithm for Factoid Type;

20. else
Algorithm 4.2 is used for the identification of the answer type for the student question containing descriptive type question words. The question words are why and how. This algorithm identifies the answer type for the both question words why & how question.

**Algorithm 4.2: Identification of Answer Type for ‘Why & How’ Type Questions**

**Input:** Question ‘q’

**Output:** Answer types for the question

1. if (question q contain “Why” words or “How” word)
    then
2.     {
     // ‘ABC’ means anything in statement of question q
4.         then
5.             Go to Answer Extraction Algorithm for Factoid Type
6.       // Rules for patterns: -
7.         How_1 == “How + much + ABC”
8.         How_2 == “How + many + ABC”;
9.         How_3 == “How + long + ABC”;
10.        How_4 == “How + for + ABC”;
11.        How_5 == “How + (Adj /Adverb) + ABC”;
12.        else if (question q == “How + ABC” && question q
Algorithm 4.3 is used for the identification of the answer type for the student question containing Factoid type question words and the simple question posed by the student for the answers. The question words are who, when, where, which, whose and whom. This algorithm identifies the answer type for who, when, where, which, whose and whom question words containing question.

**Algorithm 4.3: Identification of Answer Type for other ‘WH question words’ and ‘other simple questions except for question words’**

**Input:** Question ‘q’

**Output:** Answer types for the question

1. If (question q contain words “Who / When / Where / Which / Whose / Whom + ABC”)
2. then
3. {
4. Go to Answer Extraction Algorithm for Factoid Type;
5. }
6.
7. // ‘ABC’ means anything in the question q
8. else if (question q = = “Who + aux + (Person Name))

   Then
Chapter 4. Multimedia and Agent Based Question Answering System for eLearning

Question classifications are based on the words used in the student questions, e.g., named entities, keywords, grammatical forms, syntax, and semantics. Figure 4.18 shows that the question has several main classes: “WH-TYPE,” “non-WH-TYPE,” “Scaling.” Classification based on question words is presented in Figure 4.19. WH-TYPE questions are: who, whom, what, when, where, why, which, how, and whose type sample question as shown in Table 4.14. Question Types are:

There are four types of questions that are used in MAQAS System. For identifying the type of answers of the respective questions, different techniques have been developed.
Table 4.14: Classification and Question Focus of Different Question Words or Phrases

<table>
<thead>
<tr>
<th>Sr.</th>
<th>Question</th>
<th>Usage</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Who</td>
<td>To ask about person</td>
<td>Who opened the window?</td>
</tr>
<tr>
<td>2</td>
<td>Where</td>
<td>To ask about place</td>
<td>Where is my book?</td>
</tr>
<tr>
<td>3</td>
<td>Why</td>
<td>To ask about</td>
<td>Why are you sitting here?</td>
</tr>
<tr>
<td>4</td>
<td>When</td>
<td>To ask about time</td>
<td>When did he leave the class?</td>
</tr>
<tr>
<td>5</td>
<td>How</td>
<td>To ask about manner</td>
<td>How was your exam?</td>
</tr>
<tr>
<td>6</td>
<td>What</td>
<td>To ask about object, idea or action</td>
<td>What is your favorite subject?</td>
</tr>
<tr>
<td>7</td>
<td>Which</td>
<td>To ask about choice</td>
<td>Which color do you want?</td>
</tr>
<tr>
<td>8</td>
<td>Whose</td>
<td>To ask about possession</td>
<td>Whose turn is it?</td>
</tr>
<tr>
<td>9</td>
<td>Whom</td>
<td>To ask about object of verb</td>
<td>Whom did you see?</td>
</tr>
<tr>
<td>10</td>
<td>What Kind</td>
<td>To ask about description</td>
<td>What kind of literature do you like?</td>
</tr>
<tr>
<td>11</td>
<td>What Time</td>
<td>To ask about time</td>
<td>What time are we going to leave today?</td>
</tr>
<tr>
<td>12</td>
<td>How Many</td>
<td>To ask about quantity</td>
<td>How many pens are there?</td>
</tr>
<tr>
<td>13</td>
<td>How Much</td>
<td>To ask about amount, price</td>
<td>How much cash do you have right now?</td>
</tr>
<tr>
<td>14</td>
<td>How Long</td>
<td>To ask about duration, length</td>
<td>How long will it take?</td>
</tr>
<tr>
<td>15</td>
<td>How Often</td>
<td>To ask about frequency</td>
<td>How often have you been here?</td>
</tr>
<tr>
<td>16</td>
<td>How Far</td>
<td>To ask about distance</td>
<td>How far is Islamabad from Karachi?</td>
</tr>
<tr>
<td>17</td>
<td>How Old</td>
<td>To ask about age</td>
<td>How old are you?</td>
</tr>
<tr>
<td>18</td>
<td>How Come</td>
<td>To ask about reason</td>
<td>How come I cant see her?</td>
</tr>
<tr>
<td>19</td>
<td>Define</td>
<td>To ask about the object, action, idea.</td>
<td>Define this sentence for him</td>
</tr>
<tr>
<td>20</td>
<td>Explain</td>
<td>To ask about description</td>
<td>Explain the lecture again</td>
</tr>
<tr>
<td>21</td>
<td>List</td>
<td>To ask about description</td>
<td>List the things first that you want to buy</td>
</tr>
<tr>
<td>22</td>
<td>Show</td>
<td>To ask about description</td>
<td>Show your result sheet</td>
</tr>
<tr>
<td>23</td>
<td>Calculate</td>
<td>Fact and figures</td>
<td>calculate the heat of formation</td>
</tr>
<tr>
<td>24</td>
<td>Do</td>
<td>To ask about Yes / No</td>
<td>Do you think she will tell me the truth?</td>
</tr>
<tr>
<td>25</td>
<td>Is</td>
<td>To ask about Yes / No</td>
<td>Is there any more cake?</td>
</tr>
<tr>
<td>26</td>
<td>Can</td>
<td>To ask about Yes / No</td>
<td>Can I ask you a question?</td>
</tr>
<tr>
<td>27</td>
<td>Would</td>
<td>To ask about Yes / No</td>
<td>Would you like to read this article?</td>
</tr>
<tr>
<td>28</td>
<td>If</td>
<td>To ask about conditional Yes / No</td>
<td>If you study hard, your English will improve</td>
</tr>
<tr>
<td>29</td>
<td>Has</td>
<td>To ask about Yes / No</td>
<td>Has anybody ever told you that you are a genius?</td>
</tr>
</tbody>
</table>
Now we discussed four types of questions, which are handled in the MAQAS system as follows:

- Definition type questions
- Description type questions
- Factoid type questions
- Image based questions

### 4.9.3 Definition Type Questions

In definition type of questions, answers are given to the students in short form. These answers consist of 3 to 4 sentences commonly. In these types of questions, students want to get the necessary information about the focusing words in the questions. Following example present the definition type of question.

- What is a keyboard?

### 4.9.4 Description Type Questions

In this description type of questions, answers are given to the students in details. These answers consist of about 3 to 5 paragraphs. In these types of questions, students want to get the detailed information about the focusing keywords in the questions. Following example describe the description type of question.

- Why is ram used in the computer system?

### 4.9.5 Factoid Type Questions

In this factoid type of questions, answers are given to the students in facts, numbers and figures or few words. These answers consist on about 1 sentence or 1 to 4 words or facts and figures. In these types of questions, students want to get the specific information or facts about the focusing keywords in the questions. Following example describe the factoid type question.

- When was the first computer invented?
4.9.6 Image based Type Questions

In the image based questions, answers are given to the students that consist of images or detailed text about image. So, the image type of answers is found in the form of web sources using the focusing title names in posted question. Following example describe the image based type of question.

- What is the use case diagram of the Pakistan air traffic control management system?

In these “Text / image” based types of questions, those answers are given to the students that consist of both text and images. So, following example describe the “text + image” based type of question.

- What are the Pakistan air traffic control management systems and draw use case diagram?

For ontology development WH type Question: Question that included the question letters “W” and “H”. Question words are called as WH questions. Other classes are Hypothetical Questions, Tag Questions, Embedded Questions, Leading Questions, and Yes/No Questions.

4.10 Answer Extraction and Processing Module

The first step in answer extraction process is to retrieve the document from the web by using different search engines. Then retrieved documents are used for answer extraction. The functions of the AEP Agent is shown in Table 4.15

4.10.1 Web Documents Retrieval

The text documents are mandatory for the answer extraction process. The relevant documents are retrieved from World Wide Web. In this section, text documents containing the relevant material are extracted from web sources. In this step, a number of documents are retrieved from the different web sources in which our required information may exist. After this, required information is extracted from the retrieved documents. So for this purpose, we use the different web sources for retrieving the documents using search engines. In which are Google, Wikipedia, and webopedia search engines are used in MAQAS.
Table 4.15: Responsibilities and Activities of AnsweringExtraction&Processing Agent

<table>
<thead>
<tr>
<th>Role: AnsweringExtraction&amp;Processing (AEPA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description: 2.7. AnswerExtractions&amp;Processing Agent receives result of the keywords and headwords extracted by the KE Agent. It also receives result of the question type and answers types identified by the QAI Agent. It used the three type of search engines to retrieve the passages containing the answer from the web sources. It also communicates with Feedbackandrelevancy Agent.</td>
</tr>
<tr>
<td>Protocols: SetLMS, SetMOQA, SetQATA</td>
</tr>
<tr>
<td>Activities: ActiveLMSfunction, ActiveMOQAfunctions, ActiveQATAfunctions</td>
</tr>
<tr>
<td>Permissions: reads acquaintances</td>
</tr>
<tr>
<td>Responsibilities and liveliness: LMSA= (SetLMS, SetMOQA, SetQATA, ActiveLMSfunction, ActiveMOQAfunctions, ActiveQATAfunctions)</td>
</tr>
<tr>
<td>Safety: A successful connection with the database and LMS is established</td>
</tr>
</tbody>
</table>

4.10.2 Google

We retrieve the documents from the number of top most web pages through the Google search engine. Since Google search is most widely used among the researchers, students, and the users. So, we used this search engine for retrieving the documents from the different web pages. The Google answering system is also integrated with MAQAS to achieve the better answers. the MAQAS does not handle a question then Google question answering system is used for the answering of the question. It is further discussed in section 4.13. Top ten urls searched from Google are shown in Figure 4.20.

4.10.3 Wikipedia

Wikipedia is an encyclopedia which is also used for the information retrieval purposes. It is a entirely free encyclopedia of information available on world wide web. Wikipedia is hosted and supported by the nonprofit organization Wikimedia. Wikipedia is content-based management system that provides the helpful information for the question answering system. In the MAQAS, Wikipedia is also used as a search engine for the text document extraction. Google search engine is also referred the Wikipedia link for the basis of the keywords searched for the question posted by the students. If the contents are already searched from the Wikipedia linked matched with Google given URL, then the Google search engine URL is eliminated from the list and another URL link is added for the retrieval of the web documents. If the URLs from Google and Wikipedia are different then data is searched for the both links.
4.10.4 Webopedia

In order to improve the performance of MAQAS system, data also retrieved the relevant information from the Webopedia web page against the question posted by the student. Since Webopedia is a dictionary for the IT professionals and students that give the basic information for different technical terms of information technology. So, we use the Webopedia in MAQAS system in order to find and extract the concise answers.

4.10.5 Answers Extraction Module

When the identification of answers type against the particular question process is completed, then this module extracts the answers of the question posted by a student. This module retrieves the relevant data and different relevant documents from the internet. Secondly, when the documents are retrieved from the web, then relevant text documents are extracted from the retrieved documents. Data for extraction from the Google and
Wikipedia are shown in Figure 4.21 and 4.22 respectively. These extracted documents will also be free from the reference links in documents. Thirdly, related answers against the particular question are extracted from these documents. If an answer does not extract from the documents, then the question will be locked for teacher answering in order to get the answer. Moreover, if the answers are extracted from the documents, then these answers are sent to the FR Agent for giving the relevance ranking to these answers.

### 4.10.6 Different Corpus Text Files

There are various corpuses used in different techniques of MAQAS system. In these corpuses, important terms are saved and these terms are related to the definition, reasoning and facts terms. So by using corpuses, the number of lines of code in MAQAS system is less. Different corpus files are used for the terms that are used in the sentence or question patterns for different algorithms of MAQAS system. These algorithms
match the terms in corpus text files to analyze further. The question Three different types of corpus files are developed and used that named as “Determiner Terms Corpus Text File, Definitions Terms Corpus Text File, and Reasoning Terms Corpus Text File. The different terms in corpus text files presented for Determiner Terms in the Corpus Text File in Table 4.16. The various terms in corpus text files presented for Definitions Terms in the Corpus Text File in Table 4.17.

The different terms in corpus text files presented for Reasoning Terms in the Corpus Text File in Table 4.18.
### Table 4.17: The Definitions Terms Corpus Text File

<table>
<thead>
<tr>
<th>Definition Terms</th>
<th>Can be defined as</th>
<th>Also known as</th>
<th>Denominate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is known as</td>
<td>Also called</td>
<td>Denominated as</td>
<td></td>
</tr>
<tr>
<td>Define</td>
<td>Named</td>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>Characterize</td>
<td>Specify</td>
<td>Means</td>
<td></td>
</tr>
<tr>
<td>Identify</td>
<td>Construe</td>
<td>Meaning of</td>
<td></td>
</tr>
<tr>
<td>Determine</td>
<td>Called</td>
<td>Delineated</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4.18: The Reasoning Terms Corpus Text File

<table>
<thead>
<tr>
<th>Reasoning Terms</th>
<th>Caused</th>
<th>Incident</th>
<th>Chance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause</td>
<td>Cause</td>
<td>Milestone</td>
<td>Episode</td>
</tr>
<tr>
<td>Reason</td>
<td>Reason</td>
<td>Phenomenon</td>
<td>Event</td>
</tr>
<tr>
<td>Accident</td>
<td>Accident</td>
<td>Proceeding</td>
<td>Experience</td>
</tr>
<tr>
<td>Adventure</td>
<td>Adventure</td>
<td>Affair</td>
<td>Go</td>
</tr>
<tr>
<td>Circumstance</td>
<td>Circumstance</td>
<td>Case</td>
<td>Occasion</td>
</tr>
<tr>
<td>Scene</td>
<td>Scene</td>
<td>Agency</td>
<td>Inference</td>
</tr>
<tr>
<td>Thing</td>
<td>Thing</td>
<td>Aim</td>
<td>Senses</td>
</tr>
<tr>
<td>Element</td>
<td>Element</td>
<td>Basis</td>
<td>Reasonableness</td>
</tr>
<tr>
<td>Matter</td>
<td>Matter</td>
<td>Causation</td>
<td>Sensibleness</td>
</tr>
<tr>
<td>Motive</td>
<td>Motive</td>
<td>Logic</td>
<td>Foundation</td>
</tr>
<tr>
<td>Purpose</td>
<td>Purpose</td>
<td>Reasoning</td>
<td>Grounds</td>
</tr>
<tr>
<td>Root</td>
<td>Root</td>
<td>Deduction</td>
<td>Motivation</td>
</tr>
<tr>
<td>Source</td>
<td>Source</td>
<td>Judgment</td>
<td></td>
</tr>
</tbody>
</table>

### 4.10.7 Extracted Text Data Documents

All textual data are retrieved from Wikipedia website and top 10 web page links searched by Google. A collection of mostly 10 to 12 documents that are retrieved from the web sources. After the cleaning process, these documents are saved with their assigned name as Doc 1, Doc 2 and so on. Hence, these set of fully extracted text documents are ready for further analysis in next module after preprocessing, which is the “Documents Processing Module”. All techniques and processes are briefly described in “Documents Processing Module”.

### 4.10.8 Remove Hyperlinks and Hash Tags

In this phase, hyperlinks and hashtags are removed from each extracted text documents Doc 1, Doc 2 and so on. Since hyperlinks and hashtags not played an important role in extracting of answers. For this purpose, versatile programming tools and logical techniques are used. So by using NLTK tool, all types of hyperlinks, hashtags, non-English and non-Urdu words or sentences are removed from each text documents that
4.10.9 Documents Processing Module

In “Documents Processing Module”, all text documents are analyzed so that proper answers can be extracted using answers extraction algorithms in which different patterns and techniques have been defined. In order to analysis each extracted text document, there are 5 components have been developed which are (1) “Sentence Tokenization Process”, (2) Training Process for POS Tagging, (3) POS Tagging Process, (4) Parsing Sentences to each extracted text document and (5) Finding Named Entities (NE). Different tools and techniques in these components are described in details.

4.10.10 Sentence Tokenization Process

In this component, all sentences in each extracted text documents are tokenized in words. Since it finds those sentences that have the same keywords matched with question keywords in order to extract relevant answers from extracted text documents. So for this purpose, NLTK Tokenization tool has been used in the MAQAS system. By using NLTK Tokenization tool, all sentences in each extracted text documents are tokenized in words. For example, a sentence in Figure 4.23 has been taken from the “Introduction to Computer.txt” document.

Now, a given sentence in Figure 4.23 is in tokenized form that is depicted in Figure 4.24.

4.10.11 Training Process for Part of Speech (POS) Tagging

In this phase prior to Part of Speech tagging to each sentence of extracted text documents, training is given to NLTK POS tagging tool in the provided text documents. In which NLTK POS tagging tool is trained in order to achieve the part of speech tagging to each sentence correctly. Then NLTK POS tagging tool is applied on each extracted text document. So that each sentence is tagged into different parts of speech such as determiners(DT), verbs(VBZ), coordinating conjunction(CC), nouns(NN), adjectives(JJ) and adverbs(ADVB).

4.10.12 Part of Speech (POS) Tagging

Part of Speech (POS) tagging is a process in which the words in sentences are classified into a set of part of speech classes such as nouns, verbs, adverbs, adjectives, determiners and the articles etc. For this purpose, a number of programming tools have been developed to tag each word in a sentence into their parts of speech. So NLTK tool has
A computer system is one that is able to take a set of inputs, process the input data and create a set of outputs. This is done by a combination of hardware and software.

The diagram below shows you the idea of a computer system in its most basic form.

The computer system has one or more inputs to provide data. This data is then processed in some way. The outcome of the processing is sent to an output or it may be stored until some event happens to cause it to be output. For processing to take place, there needs to be a set of instructions or what needs to be done. This set of instructions is called a program. This system is called a stored-program computer.

Figure 4.21: Data Extracted from Google Search Engine

The computer program is a collection of instructions that performs a specific task when executed by a computer. A computer requires programs to function, and typically executes the program's instructions in a central processing unit.

A computer program is usually written by a computer programmer in a programming language. From the program in its human-readable form of source code, a compiler can derive machine code—a form consisting of instructions that the computer can directly execute. Alternatively, a computer program may be executed with the aid of an interpreter.

Figure 4.22: Data Extracted from Wikipedia Search Engine
been used in our MAQAS System to tag each word in sentences of extracted documents into parts of speech. Now all words in each sentence of each extracted text document are tagged into parts of speech. For example, Figure 4.24 is depicted the POS tagging to each word of a given sentence in Figure 4.23.

4.10.13 Parsing Sentences to each Extracted Text Document

Parsing is a process that constructs the syntactic structure over a given sentence which is defined by the formal grammar. Some previous parsers tools do not give the correct result and need the set of grammatical rules for parsing the sentence. However, recently developed parser tools are better and advance to parse the sentence directly by using complicated statistical models. Now, NLTK parser is an advanced and better parser tool that has been used in MAQAS system of this research. In this phase, all sentences of each extracted text document are parsed in order to know the syntactical structure of sentences. Since syntactical structure describes that how words are correlated to each other within a sentence. For example, Figure 4.24 is depicted the parsed syntactical structure of a given sentence.

4.10.14 Finding Named Entities (NE)

Named Entity (NE) tool is a named entity recognition tool that identifies and extracts the entities from the textual data. It is the subtask of Information Extraction that identifies the named entities and classifies them into pre-defined classes. Different types of classes or categories have defined for named entities such as names of person, companies, organizations, locations, time, date, quantities, percentages and financial values etc. In this component, named entities from all sentences in each extracted text document are extracted. Many programming tools are available to find the Named Entities in textual
documents. But in MAQAS system, AlchemyAPI Named Entity tool and Stanford N-ER Tagger tool has been used. Extracting of Named Entities in this MAQAS system has an important role for extracting the factoid type answers from the text documents. However, Figure 4.26 presented that sentence extracted from the text document. For example, following named entities with their named entity category have been found by the AlchemyAPI named entity tool that is depicted in Figure 4.27.

### 4.10.15 Named Entities Categories

<table>
<thead>
<tr>
<th>Named Entities</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luigi Menabrea</td>
<td>Person</td>
</tr>
<tr>
<td>Boston</td>
<td>City</td>
</tr>
<tr>
<td>software developer</td>
<td>JobTitle</td>
</tr>
<tr>
<td>software systems</td>
<td>FieldTerminology</td>
</tr>
<tr>
<td>object code</td>
<td>FieldTerminology</td>
</tr>
<tr>
<td>scripting languages</td>
<td>FieldTerminology</td>
</tr>
<tr>
<td>virtual machine</td>
<td>FieldTerminology</td>
</tr>
<tr>
<td>Knuth</td>
<td>Person</td>
</tr>
<tr>
<td>Fasculines</td>
<td>Company</td>
</tr>
<tr>
<td>non-volatile memory</td>
<td>FieldTerminology</td>
</tr>
<tr>
<td>Manchester</td>
<td>City</td>
</tr>
<tr>
<td>Microsoft</td>
<td>Company</td>
</tr>
</tbody>
</table>

**Figure 4.25:** Named Entities Identified by Entity Extractor Tool

Now, Named Entities have been found for the following sentence in

Figure 4.25 by the Stanford NER Tagger that classifies the named entities into seven classes or categories.

Figure 4.27 is depicted the named entities with their respective classes identified by Stanford NER Tagger. Moreover, ‘O’ term means “Other” which describes that particular word in not a named entity and belongs to other categories in Stanford NER Tagger.

This machine is considered by some to be the origin of the stored-program computer—used by John von Neumann (1946) for the "Electronic Computing Instrument" that now bears the von Neumann architecture name.

---

**Figure 4.26:** Extracted Sentence for Answer
4.11 Extracting Keywords and Relevancy Score Module

When the task of analysis of extracted text documents has been completed, then extracting of keywords from each document process is started. In this module keywords from each text document are extracted and given the relevance score to each keyword. So, “Extracting Keywords and Relevancy Score Module” has two components which are one is (1) “Extracting Keywords” component and other is (2) “Finding Relevancy Score of Keywords Techniques” component. Tool and technologies in these two components are described briefly.

4.11.1 Extracting Keywords

Keywords play an important role in QA framework in order to find the related answers from textual documents. So, in order to extract the keyword from textual data, the particular keyword extraction algorithms adopt the following procedures:

- Exclude any punctuation marks and stop words from the given textual data.
- Extract all the possible words and phrases that can be the possible keywords.

So these extracted words are the candidate keywords.
To calculate the properties for each candidate keyword which identifies that it may be a characteristic keyword. The different properties for selection of typical keywords are the frequency of occurrences for a specific word, position in the sentence and document, the length of phrasing words, similarity to other candidate keywords, Is it a popular word?, and the length of sentences.

Figure 4.29 describes how characteristic keywords are extracted from the base of their properties commonly by a number of keywords extraction tools. In the MAQAS system, keywords are extracted by using AlchemyAPI Keywords Extractor tool. Now keywords for each extracted text document are extracted and sent to the next component for further processing.

4.11.2 Finding Relevancy Score of Keywords Techniques

In this phase, the relevance score of each keyword for each text document is found. Since, relevance scoring to each keyword plays an important role in extracting the relevant and concise answers from text documents. In MAQAS system, relevancy score has also been given to keywords through AlchemyAPI Keywords Extraction tool.

4.11.3 Answers Extraction

Then keywords are extracted from the question. Now from the keywords in the given question, headwords are identified. Those are the major focus of the question. Then the answer is extracted on the basis of these headwords. Then the answer is processed according to the headwords. If the answer is already extracted and stored in the repository, then it is retrieved from the knowledge repository and delivered to the student. In case if the answer does not exist in the knowledge base then related documents are extracted from the web. Web documents may be in the form of PDF, DOC, PPT. Then paragraphs are extracted from the selected documents. These paragraphs are synthesized to the answer of the given question. Another case, if the system fails to find the answer to the question the system preserves the question. This question is locked by the system for the further answering by the instructor.

In this process, relevant answers are extracted from the retrieved documents. For extracting the answers, different patterns and techniques have been developed. When the answers are extracted from the set of retrieved documents, then it is sent to the next module that is FR Agent for getting relevancy scores and feedback about the extracted
Figure 4.28: Answer Extraction and Processing Module
answers from the students. Moreover, if the answer to the posted question is not extracted from the set of retrieved documents, then that question is locked for getting the answers from the teacher.

When the task of finding the relevancy score of keywords is completed, then information about these keywords and their relevance scores are sent to the “Answers Extraction Module”. In this module concise and proper answers are extracted from the set of extracted text documents. So, “Answers Extraction Module” as shown in Figure 4.28, consists of main two components which are (1) “Answers Extraction Algorithms” and other is (2) “Answers Ranking Techniques”. Moreover, there are two sub-components that are (1) Extracted Answers from extracted Text documents and (2) Candidate Answers. So, the syntactical and logical techniques have been adopted in these components for extracting the concise answers that are related to the posted question in which user asking for the focus keyword. Techniques and procedures are briefly discussed.

4.11.4 Answers Extraction Algorithms

In this component, precise and relevant answers are extracted against to the question posed by the user through different techniques. In these techniques, different types of patterns and corpus text file have been made. By using these patterns, tokenized, parsed
and POS tagging sentences are analyzed. So, in order to extract the concise and related answers from the text documents, there are three types of algorithms are developed in which different patterns and techniques have been adopted. These three algorithms are (1) “Definition Type Answers Extraction Algorithm”, (2) “Description Type Answers Extraction Algorithm” and (3) “Factoid Type Answers Extraction Algorithm”. However, before to proceeds these algorithms, information about answers type detection are needed about the posed question. So that one type of answers extraction algorithm may run for extracting the concise answers. Since answer type detection for the posed question has already determined in “Natural Language Question Processing Module”. Now it can proceed and perform the syntactical task on sentences from the text documents through “Answers Extraction Algorithms”. Three types of answers extraction algorithms are presented in which all steps are described to find the most suitable answers from the set of text documents. Algorithm 4.4 is developed for those questions that have definition type of answers.

Algorithm 4.4: Answer Extraction Algorithm for Definition Type Questions

**Input:** a) Question ‘q’

b) Set of all extracted text documents ‘TD’

**Output:** Definition type Answer for the given question

1. Start // Sentence Patterns

2. 

3. ‘ABC’ means anything in the sentence ‘s’ of respective document ‘td’

4. Sp1: \{ ABC + (“called”/ “named as”) + [ Noun / (JJ+Noun)] \}

5. Sp2: \{ [ Noun / (Adj+Noun)] + (Definition Terms Corpus Text File / “can be defined as”) + ABC \}


7. Sp4: \{ ABC + (“is known as”/ Definition Terms Corpus Text File) + [ Noun / (JJ+Noun)] \}

8. Sp5: \{ [Noun / (JJ+Noun)] + aux + ABC \}

9. 

10. if (sentence patterns \(=\) “Sp1 / Sp2 / Sp3 / Sp4 / Sp5”) then
11. return it as a candidate answer
12. else
13. return “Document does not contain the answer”
14. end if;
15. end;

Algorithm 4.5 is developed for those questions that have factoid type of answers. These questions are commonly used the question words, “where” for the answer type is the place. While in the case of question, the word is when then answer type is time or duration. The other factoid type answer for the question words are “who, whose, whom, how long, how many and, etc.

Algorithm 4.5: Answer Extraction Algorithm for Factoid Type Questions

**Input:** a) Question ‘q’
b) Set of all extracted text documents ‘TD’

**Output:** Definition type Answer for the given questions

1. Start // Headword is the first [Noun / Noun Phrase] after question word
2. Extract keywords ‘k’ from question ‘q’
3. Extract headword ‘h’ from question ‘q’
4. if (question’s keywords ‘k’ exist in the sentences ‘s’ of text document ‘td’) then
5. {
6. Extracting those sentences ‘s’ that having question’s keywords ‘k’
7. Extracting the ‘Named Entities (NE)’ from document ‘td’
8. question_word = = Extract Question word ‘Qw’ from question ‘q’
9. if (‘Qw’ of question q = = “who / whose / whom”) then
10. Extract that word ‘w’ having the Named Entity (NE) type of “Name/Person”;
12. return it as a candidate answer
13. else if (‘Qw’ of question q = “When”)
14. then
15. Extract that word ‘w’ having the Named Entity (NE) type of “Date/Duration/Time”;  
16. return it as a candidate answer
17. else if (‘Qw’ of question q = “Where”)
18. then
19. Extract that word ‘w’ having the Named Entity (NE) type of “Place/Location Name”;  
20. return it as a candidate answer
21. else if (‘Qw’ of question q = “Which”)
22. then
23. Extract that word ‘w’ having the Named Entity (NE) type of “Object / Organization Name”  
24. return it as a candidate answer
25. else if (‘Qw’ of question q = “How much / How many / How for / How long / (How + Adj/Adv)”)
26. then
27. Extract that word ‘w’ having the Named Entity (NE) type of “Number / Degree”  
28. return it as a candidate answer
29. end if;
30. end;

Moreover, Algorithm 4.5 is developed for those questions that have the factoid type of answers such as Person name, any organization or company name, data and time. Keywords and headwords are focusing words of a respective question. In which headwords
are those words that come first after the question words in questions. So, different patterns are developed in order to extract the factoid type of answers which are shown in Algorithm 4.5. Algorithm 4.6 is developed for those questions that have the description type of answers. So, different patterns are developed in order to extract the description type of answers which are shown in Algorithm 4.6.

**Algorithm 4.6: Answer Extraction Algorithm for Description Type Questions**

**Input:**
- a) Question ‘q’
- b) Set of all extracted text documents ‘TD’

**Output:** Descriptive type Answer for the given question

1. Start
2. Extract ‘keywords’ from question ‘q’
3. if (question’s ‘keywords’ are contained in the sentences ‘s’ of document ‘td’)
4. then
5. Extract those ‘paragraphs’ that having question’s ‘keywords’ in sentence ‘s’
6. return it as candidate answers
7. Else
8. return “Document does not contain the answer”
9. end if;
10. end;

**4.11.5 Extracted Answers from Extracted Text Documents**

When the most related answers are extracted from the extracted text documents through the Answers Extraction Algorithms 4.4, 4.5 and 4.6, then these answers will be the relevant answers to the posed question. And these answers are the extracted answers. But these answers are not arranged in ranked wise. For this purpose, we have developed the new techniques for assigning the rank score to extracted answers on the base of their high priorities. So, extracted answers are referred to the next component that is “Answers Ranking Techniques”.
### 4.11.6 Answers Ranking Techniques

Extracted answers are extracted from the different text documents and not arrange in an ordered list according to their given rank score. Because it is the most probably chance of that most related answers may be presenting at the low position than other answers in the list of candidate answers. FR agent handled the extracted answer relevancy. FR Agent activities are shown Table 4.19. In which relevance scores of all the keywords are added to each extracted answer separately. Secondly, cosine similarity is calculated for each extracted text document with a question. So for calculating the cosine similarity between the question and each text document separately, the TF (Term Frequency) score and the IDF (Inverse Document Frequency) score for the keywords in each text document are needed.

TF calculation is a process in which each term in the document gives the score due to its number of times occurrences in the particular text document.

| **Table 4.19:** Responsibilities and activities of FeedbackRelevancy Agent |
|------------------|--------------------------------------------------------------------------------|
| **Role:** FeedbackRelevancyAgent (FRA)                                                                 |
| **Description:** This agent is responsible for establishing a connection with student feedback score module and automatically calculated relevancy score of the answer extracted from the web sources and delivered to student. It also activated some function of appointment fixing agent. |
| **Protocols:** SetLMS, SetMOQA, SetFRA                                                                 |
| **Activities:** ActiveLMSfunction, ActiveMOQAfunctions, ActiveFRAfunctions                            |
| **Permissions:** reads acquaintances                                                                     |
| **Responsibilities and liveliness:** LMSA= (SetLMS/SetMOQA /ActiveLMSfunction/ ActiveMOQAfunctions/ActiveFRAfunctions) |
| **Safety:** A successful connection with the database and LMS is established                            |

The IDF is a technique which calculates that how many times a unique keyword occurs in different documents. Moreover, cosine similarity gives the information about the similarity between the two documents. So these three mechanisms have been used for finding the cosine similarity between the question and each extracted text document separately. The following equations describe that how the Term Frequency (TF), the Inverse Document Frequency (IDF) and the Cosine Similarity are calculated.

Term Frequency value of keywords by using the following Equation 4.1.

\[
TF = \frac{F}{N} \quad (4.1)
\]

Where,
‘F’ describes the number of times occurrences for a specific keyword in the particular text document
‘N’ describes the total number of words in the particular text document
To find the Inverse Document Frequency value of the particular term as shown in Equation 4.2.

\[
IDF = 1 + \log_e((\text{Total Number of Documents})(\text{Documents containing specific term }))
\]

(4.2)

For example, if there are total five documents and term “Computer” occurs in three documents. Then IDF value for “Computer” term is calculated by following:

\[
IDF(\text{Computer}) = 1 + \log_e(5/3)
\]

\[
= 1.510526
\]

Since in the MAQAS system, cosine similarity is calculated between the question and each extracted document separately. To find the cosine similarity score between the question and document, take the dot product of TF from document and TF from the question that is multiplied by IDF. So this cosine similarity equation between the question and document is completely defined in the following Equation 4.3.

Now, Cosine Similarity Equation 4.3 is:

\[
\text{Cosine Similarity (Q,D)} = (\{(\text{TF from query} \times \text{IDF})(\text{TF from Document})\})(\text{TF query} \times \text{IDF} \times \text{TF Document})...(4.3)
\]

Where, “Q” is the question
“D” is the text document
“TF” from question means that number of times occurrences of a specific term in the question
“TF” from document means that a number of times occurrences of a specific term in the particular text document
“IDF” is the inverse document frequency

The values for each document are being calculated using the cosine similarity formula separately. Moreover, the total relevance score for each extracted answer is calculated
by adding the relevance scores of all the keywords that occur in their respective answer sentence. Then those answers will be given the high priorities which are extracted from those text documents that have the high cosine similarity value.

It means that if a text document has a high cosine similarity value to the question, then extracted answers from this document will be at the top position in the ranked list of extracted answers. When this task is complete, then extracted answers will be arranged their total relevance scores accordingly. Sometimes more than two answers are extracted from the same document, then these answers are arranged in the top to bottom order that has the high relevance score values of their sentences. So, Algorithm 4.7 is developed that is named as “Ranking to Extracted Answers” algorithm in which all the procedures and steps to give the ranks of answers are described.

**Algorithm 4.7: Ranking of Extracted Answers**

**Input:**

a) Question ‘q’

b) Set of all extracted text documents ‘TD’

c) Set of all answers extracted by answers extraction algorithms

**Output:** Answers will be ranked according to their relevancy in ranked list.

1. Text_documents_set[td] “Set of all extracted text documents ‘TD’

2. for (each document td in documents_set[td])

3. do{

4. for (each word w in documents td)

5. do {

6. Tf = = “count the each term occurs in a number of times in the document td”

7. N = = “count the total number of terms in document td”

8. \( \text{TF}_D = \frac{F}{N} \)

9. Total_docs = = “total number of extracted text documents”

10. L = = “total number of documents which have the same term w in it”

11. \( \text{IDF} = 1 + \log_e(\text{Total_docs}/L) \);
for (each word w in question q)

Do {

FQ = “count each term occurs in a number of times
in the question q”

NQ = “count the total number of terms in question q”

TF_Q = FQ / NQ

M = TF_Q * IDF

TF_DQ = “get TF_D score of question’s keyword”

R = 0; S = 0

R = R + ( M * M )

S = S + (TF_DQ * TF_DQ)
}

end for

similarity btw q and d == (cosine(MTF_DQ)) / (sqrt(R)*sqrt(S))

end for

if (text document td have high similarity score than other documents[td]

&& answers are also extracted from this document td)

Then

{ 

Arrange answers according to their high total score of keywords;

Display arranged answers of this document td;

Remove this document from documents [td] list;

}

end for

end
4.11.7 Candidate Answers

Extracted answers are arranged in ranked list through the Algorithm 4.7, then these answers will be the candidate answers. In which most suitable related answers are presented at the top position, and other low ranked candidate answers are presented at the bottom position in ranking wise. So, these candidate answers are concise answers and are the most related answers to the focused keywords of the user question that is asked by the user.

4.12 Image based Answers Module (IA Agent )

In this module, Answers are extracted for the image type questions. For this purpose, IA Agent s the title or image name from the question posted by a student. Then this image name or title is searched from the set of images of XML formats. Image name also searched from the crowdsource engineering vocabulary also containing other information.

If the image name or title is mapped to the name or titles of XML format images, then extract these images and provide information about these images as shown in Figure 4.30. In this MAQAS system images are searched for the type of Use Case Diagrams and Books. These Use Case Diagram or Books images types are extracted from the XML images set and delivered to the student.

4.12.1 Image question Processing

The annotator application is developed for image annotation. The annotator module annotates the images and also generates XML files as shown in figure 4.31. This application also validates this XML file. When any diagram is entered into the application, it is annotated and its XML file generated using crowd-source engineering. Fig 7 depicts the use case diagram and the XML file generated.

UML Image Annotation and Validation Algorithm

The XML file is by using the crowd-sourced engineering. The validation is also shown in the right tab of the tool. The UML image annotation and validation algorithm are shown in the Algorithm 4.8.

Algorithm 4.8: UML Image Annotation and Validation
Figure 4.30: Image base Question Processing for Answers
Chapter 4. *Multimedia and Agent Based Question Answering System for eLearning*

**Figure 4.31: Image Annotator Tool for the Annotation and Validation of Images**

**Inputs:** Image files Imgi, where I = 0, 1, 2, ..., n

**Output:** Annotated XML files

1 Start
2 Take Imgi, Image and load it into image box
3 Loop I = 0, 1, 2, ..., n
4 Check the current folder
5 If (ImgiXML = No)
6 Load it in XML Editor
7 Else
8 Load a template “XML” file.
9 Save file with “Imgi.xml” extension
10 End IF
11 Check the current folder
12 If (ImgiXML) = Yes
13 Load it in XML Parser
14 Else
15 Load a template “XML” file.
16 Save file with “.xml” extension
17 End if
18 End loop
19 End

In this module, get the relevancy score and also get feedback from the students for the extracted answers. If the feedback is “Excellent” and relevancy score is “Good” from the student, then this extracted answer is correct. Moreover, this correct answer is delivered to the student as shown in figure 4.32.

If the feedback is “Satisfactory” and relevancy score is “Greater than 60 %” from the student, then this extracted answer is also correct. And this correct answer is delivered to the student.

4.12.2 Student Ranking Answer

The automatic answers are extracted by the tool and delivered to the students. Now students provide the feedback about the delivered answer and ranked them in the among three levels. These three levels are:

- Excellent
- Satisfactory
- Unsatisfactory

If the feedback is “Unsatisfactory” and relevancy score is “Good” from the student, then this answer is not correct. Moreover, fix the appointment of teacher for the student in order to give the more suitable answers from the teacher using the multimedia devices.

4.12.3 GPS Coordinates Collection

Multimedia based data GPS coordinates collection for the intelligent appointment system. In “Appointment Fixing” function, firstly get the student GPS location by using Google Map. If the student is free and exist in the no noisy place such as Home, Office,
Figure 4.32: Answer Delivery and Appointment Module
and Class Room. Moreover, teacher slot is not free, then fix the available appointment for giving the answer to the student using available synchronization multimedia devices on both ends. If the student is free and but exist in the noisy place such as Traffic area, Public area, and Market area. Then, the AF Agent collaborated to the student for their availability for giving the answer to the student.

4.12.4 Google QA System

The Google answering system is also integrated with MAQAS to achieve the better answers. If the MAQAS does not handle a question then Google question answering system is used for the answering of the question. Answer snippet generated by the Google is extracted and delivered to the student in case of meet the relevancy. Figure 4.33 represents the answer extracted from the Google snippets for delivery to the students.

![Multimedia & Agent based Question Answering System (MAQAS)](image)

**Figure 4.33:** Answer Extracted from Google QA Snippets

4.12.5 User Questions and Extracted Answers Repository

The repository is a database that plays an important role as a central database in the network in which keep data and maintain it in an organized form. A repository is a place where particular databases, files or data are fetched for further distribution or transfer over the network. It can be directly accessible to the users depending on the different conditions across a network. In this phase, extracted candidate answers of the particular questions are stored in a repository database. In this component, reformulated
questions that were posed by the user are stored on the online repository database along with their extracted particular candidate answers. So that, whenever the user enters the same question in the MAQAS system, then this question will be checked from the online repository database. If the question matches to any questions in the record that are stored in the database, then candidate answers this question will be retrieved from the database. Moreover, if the question is not match to any stored questions in the database, then the answers to this question will be found from the web sources and all the tasks of the are performed for extracting the most related answers for the user.

In order to improve the accuracy performance and for developing the efficient MAQAS system, the “Zoho Creator” database has been chosen as a repository. “Zoho Creator” is a database, in which we create a new online database. It provides the easy interface for creating the new database by using drag and drop features easily. To create a database on “Zoho Creator”, it gives the facility to define our own rules for processing the data. It also presents the information about the users and gives the information that which time and how many times the specific questions are created and accessed. Anyhow, it is good and most used online database in which any user can create their own databases quickly. So, a new database has been created on “Zoho Creator” with the name of “MAQAS system” that can be accessed by the following link through the MAQAS system.

Since, when the question enters by the user in the MAQAS system, then the link has already been embedded with login details in the source code for accessing the MAQAS database. So that transaction could be easily handled and maintained without any interruptions. The two fields are made in the MAQAS database with the name of “User Questions” and “Candidate Answers”. In User Questions field all the reformulated questions will be stored, and in the “Candidate Answers” field all the set of extracted answers against the particular questions will be stored.

When a Student enters the question into “MAQAS system”; then the task for the reformulation process of the question in the natural language question processing module is performed. Then after this reformulation process, a reformulated question is First sent to the online repository database of MAQAS system. Now, when the question is entered in the database; then a process is started for matching this question with other questions that already exist in the database.

If the posted question matches with the current questions in the “User Questions” field,
Figure 4.34: Repository for MAQAS System

Figure 4.35: Answer for Repository of MAQAS System
then the respective candidate answers will be retrieved from the record of “Candidate Answers” field against this question as shown in the figure 4.35.

If the incoming question is not match with other saved questions in the database, then this new incoming question will be stored firstly in the record of the “User Questions”, and then this question will be sent to other components and next modules for further processing. After this when the all syntactical analysis processes in the modules of MAQAS system are completed, then the candidate answers will be sent to online repository database MAQAS system and saved in the record of “Candidate Answers” field against this question. Figure 4.34 shows the repository of MAQAS system.

4.13 Case Study

Let’s a simple query “what is the role of the operating system?” run on MAQAS system and take a subset of the keywords and calculate the relevance for the decision-making at a later stage to fix appointments of the student-teacher interaction in synchronous communications.

4.13.1 Keywords Relevancy in Answers Response to Student Question

Table 4.23 denoted the keywords relevancy on small subset keywords of the answer delivered to the student that is calculated by using the Equation (4.4).

\[ Wi = (TF_i)(IDF_i) = tf\log\frac{No.\ of\ Doc}{df_i} \] ..........................E(4.4)

Where “\(TF\)”denotes Term Frequency in Question

“\(IDF\)”stands for Inverse Document Frequency

“\(tf\)”stands for in Term Frequency in Text Document

“\(df\)”stands for document frequency containing specific term.

In Table 4.20 keyword’s relevancy is measured by using the different formulae and algorithms. The answer extracted from the web sources and documents. These extracted documents are used to further extracting the relevance text to compose the answer to the student question. The text is extracted along with the relevance values of the terms. The answers are extracted more one from the collection of the documents. These extracted answers for each question are ranked according to the relevance of the terms containing.
### Chapter 4. Multimedia and Agent Based Question Answering System for eLearning

#### Table 4.20: Keywords Relevancy Calculations in Answers to Student

<table>
<thead>
<tr>
<th>Text</th>
<th>Type</th>
<th>Count</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application software</td>
<td>Field Terminology</td>
<td>13</td>
<td>0.784717</td>
</tr>
<tr>
<td>Computer programmer</td>
<td>Field Terminology</td>
<td>3</td>
<td>0.772240</td>
</tr>
<tr>
<td>Operating system</td>
<td>Field Terminology</td>
<td>14</td>
<td>0.713496</td>
</tr>
<tr>
<td>Programming language</td>
<td>Field Terminology</td>
<td>9</td>
<td>0.631108</td>
</tr>
<tr>
<td>Source code</td>
<td>Field Terminology</td>
<td>7</td>
<td>0.522889</td>
</tr>
<tr>
<td>Cpu</td>
<td>Technology</td>
<td>3</td>
<td>0.423652</td>
</tr>
<tr>
<td>Analytical engine</td>
<td>Field Terminology</td>
<td>4</td>
<td>0.418333</td>
</tr>
<tr>
<td>Software development</td>
<td>Field Terminology</td>
<td>2</td>
<td>0.395880</td>
</tr>
<tr>
<td>System software</td>
<td>Field Terminology</td>
<td>4</td>
<td>0.383269</td>
</tr>
<tr>
<td>Programmer</td>
<td>Job Title</td>
<td>4</td>
<td>0.370161</td>
</tr>
</tbody>
</table>

The most relevance answer is delivered to students in response to their question. The students were after receiving the answer will provide the feedback to the system.

Now FR agent collects the feedback from the students. This feedback along with question relevancy used to measure the need for multimedia-based communications. If the student feedback is satisfactory then the AF agent will fix major role-plays in the next step towards appointment. After taking the decision of appointment will be fixed.

#### 4.13.2 Students Collected Data through Smartphone GPS

AF Agent collects the coordinates through smart-phone GPS sensor as shown in Table 4.21. These coordinates are helpful to detecting the location of the student. Once the student location is identified then AF agent decides for either immediate appointment is possible in case teacher has a free slot. Otherwise, appointment deferred to some appropriate time.

Appropriate time will have sought out by a collaboration of the student and AF agent. While AF agent also communicate to the teacher for the available free slots. the activities and responsibilities of AF is shown in Table 4.22. Therefore, the student teacher synchronous interaction is possible for a better solution of the student problems.

Now Google Map is used to find out the locations of students as shown in Table 4.21 and Figure 4.36. User’s smartphones are used to collect their Global Positioning System (GPS) coordinates. Google Maps API has used the location into coordinates like latitude and longitude dynamically, and these coordinates are used to place markers on the map as shown in Figure 4.35. Multiple markers show locations of students, whether they are in the campus or somewhere else while using the framework which is based on
Table 4.21: A Subset of Students Collected Data through Smartphone GPS

<table>
<thead>
<tr>
<th>Student #</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33.63978484</td>
<td>73.07521115</td>
<td>Campus</td>
</tr>
<tr>
<td>2</td>
<td>33.63920424</td>
<td>73.07149337</td>
<td>Swedish Institute of Technology</td>
</tr>
<tr>
<td>3</td>
<td>33.63977484</td>
<td>73.07529315</td>
<td>Campus</td>
</tr>
<tr>
<td>4</td>
<td>33.63976484</td>
<td>73.07524415</td>
<td>Campus</td>
</tr>
<tr>
<td>5</td>
<td>33.63804305</td>
<td>73.07392538</td>
<td>Lasania</td>
</tr>
<tr>
<td>6</td>
<td>33.63893521</td>
<td>73.07305366</td>
<td>Chaman Ice Cream</td>
</tr>
<tr>
<td>7</td>
<td>33.63975484</td>
<td>73.07528515</td>
<td>Campus</td>
</tr>
<tr>
<td>8</td>
<td>33.63974484</td>
<td>73.07526615</td>
<td>Campus</td>
</tr>
<tr>
<td>9</td>
<td>33.63920424</td>
<td>73.07139337</td>
<td>Swedish Institute of Technology</td>
</tr>
<tr>
<td>10</td>
<td>33.63972484</td>
<td>73.07524815</td>
<td>Campus</td>
</tr>
<tr>
<td>11</td>
<td>33.63977484</td>
<td>73.07530215</td>
<td>Campus</td>
</tr>
<tr>
<td>12</td>
<td>33.63894521</td>
<td>73.07315366</td>
<td>Chaman Ice Cream</td>
</tr>
<tr>
<td>13</td>
<td>33.64044581</td>
<td>73.0749473</td>
<td>Office, Workspace</td>
</tr>
<tr>
<td>14</td>
<td>33.63973484</td>
<td>73.07530715</td>
<td>Campus</td>
</tr>
<tr>
<td>15</td>
<td>33.63804305</td>
<td>73.07382538</td>
<td>Lasania</td>
</tr>
<tr>
<td>16</td>
<td>33.63971484</td>
<td>73.07513715</td>
<td>Campus</td>
</tr>
<tr>
<td>17</td>
<td>33.63970484</td>
<td>73.07512615</td>
<td>Campus</td>
</tr>
</tbody>
</table>

Table 4.22: Responsibilities and Activities of AppointmentFixing Agent

<table>
<thead>
<tr>
<th>Role:</th>
<th>AppointmentFixing Agent (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>Appointment Agent receives the result from the FeedbackandRelevancy Agent. Mobile It also collect location coordinates from the GPS to pinpoint the student location. To keep in view teacher free slots, the appointment agent fix the appointment of the student and teacher for synchronous communication.</td>
</tr>
<tr>
<td>Protocols:</td>
<td>SetLMS, SetMOQA, SetAFA</td>
</tr>
<tr>
<td>Activities:</td>
<td>ActiveLMSfunction, ActiveMOQAfunctions, ActiveAFfunctions</td>
</tr>
<tr>
<td>Permissions:</td>
<td>reads acquaintances</td>
</tr>
<tr>
<td>Responsibilities and liveliness:</td>
<td>LMSA= (SetLMS, /SetMOQA, /ActiveLMSfunction, /ActiveMOQAfunctions, /ActiveAFfunctions)</td>
</tr>
<tr>
<td>Safety:</td>
<td>A successful connection with the database and LMS is established</td>
</tr>
</tbody>
</table>

4.14 Appointment Fixing Module

FR Agent calculates the relevance of the answer and waits for the feedback from the students. There are some possible scenarios exist as follows:

Case 1: After the answer relevancy is calculated then the FR agent waits for the student feedback. Student ranked the answer after reading it. The answer contents delivered to the students contain the material regarding the question. If the student feels, satisfaction ranked good and relevancy of the answer is also measured as higher than the threshold,
then AF agent takes no action. As AF agent, received no input from the FR agent, hence the student-learning process is satisfactory in subject learning.

FR agent calculates the relevance of the answer and waits for the feedback from the students. There is some possible scenario exists as follows:

Case 2: If the measured relevancy is less than the threshold, and the contents are not related to the student question. Then answer is not relevant the question extracted keywords and headword. The student has no knowledge of the concept and may feel satisfactory. As the content delivered are understandable and useful then students ranked the answer very good. In this case the AF agent will remain motionless. The student learning is upward and no need to fix the appointment.

Case 3: In the eLearning environment the student is answered by the system developed during the research work. If the student below the mark ranks the auto-replied answer by the MAQAS and answer relevancy is measured less than the threshold, then FR Agent sends the result toward AF Agent, AF Agent will collaborate the teacher and student. AF Agent checks the location of the student by the help of the mobile GPS. The value of the coordinates accessed by the AF Agent is shown in Figure 4.31. In case if the student is at home are in office, then AF agent will check the teacher availability for the appointment. If the teacher has a free slot, then AF agent will fix the appointment with immediate effect and inform the student after taking the consent of the teacher.
4.14.1 Multimedia Answer Contents to Student

Now, the student can interact with the teacher by using the synchronous communication media tools like Skype, Team Weaver, Camtasia, for an audio video call or presentation as shown in Figure 4.37. Teacher delivers the multimedia contents to answer the student query at a satisfactory level. This will boost the learning process of the student and learn curve is enhanced. If AF Agent locates the student in the market or in some place where silence and comfort level is very low or student in a noisy environment. Then AF Agent will take action after collaboration with teacher and will fix the deferred appointment for the student-teacher multimedia-based interaction.

4.15 Conclusion

This research work enhances the student learning and analysis by the multimedia representation and content delivery. This agent-based architecture will provide the better student-teacher interaction to boost the learning process of the student.
Chapter 5

RESULTS AND DISCUSSIONS

This chapter presents the results of question answering system. The results are also compared with previous approaches for the performance evaluation. The proposed question answering system is named as Multimedia and Agent based Question Answering System (MAQAS). The result and performance of the system are analyzed by its performance measured and compared with the existing system.

5.1 Performance Evaluation

The performance of the system is measured and compared with the existing systems based upon the parameters. As there are different types of the questions that are used for the performance evaluation. Therefore, the developed question answering system is measured by a number of existing systems and the different parameters. These parameters are; the number of the correct answers, answers relevancy, overall performance. These parameters are further compared with the different type of question which related to two main type of questions: one is WH type question and the second is Non-WH type question. The WH type question is further divided based on their question word are question phrases. In case if the question phrase is WH type then Question words and phrases determine the answer types which are a factoid, descriptive and definition types. Separate results are prepared and compared for such type of question. Similar to that for Non-WH types question words are also used for the answer type investigating which are yes/ no, scale, factoid, descriptive and definition types. The data set of the question is taken on the subject of Introduction to Computing. This subject is related to computer science field and almost all the student of BSCS are mandatory to take this
course for the completion of their degree.

5.1.1 Parameters for evaluation

The number of document relevant retrieved from the Google, Wikipedia, Webopedia, PDF documents and search engines retrieved other documents. Then extraction of the precise and suitable answers to the student question from the retrieved documents. The answers to the posed questions are ranked based on the relevancy of the text of the answer. The retrieved documents are containing the hyperlinks, hashtags full stop and special symbols. The retrieved text documents are cleaned and saved using unique names such as D1, D2, D3... The cleaned documents from the hyperlinks and hashtags are named as refined documents. The process of extraction of the documents and then answer paragraphs for the correct answers are described already in the previous chapter.

In order to measure and compare the quality, reliability, efficiency, and performance may also be used. The measurement and analysis techniques are applied to the different type of question answering systems. For the current research work is evaluated and analyzed on various defined parameters. These measurements are Recall Measurement, Precision Measurement, Accuracy Score Measurement, F-measure, and Specificity. These five types of measurements have been used to evaluate the proposed MAQAS system. In each measurement, different measurable factors have been defined which are described separately.

5.1.2 Recall Measurement

Recall measurement is basically a measurement that is used to measure the performance of searching techniques. It is most widely used by many researchers in order to evaluate the searching techniques adopted in their systems. So, I used Recall Measurement in order to evaluate the proposed MAQAS system. Since Recall Measurement tells us that how much the number of extracted relevant records are related with the total number of relevant records. So, we defined Recall Measurement for the proposed MAQAS system that it is the ratio of the number of the relevant answers and the retrieved answers to the total number of relevant answers. Following the Equation 5.1 describes the Recall measurement.

Recall=\frac{TP}{TP+FN} \quad (5.1)
5.1.3 Precision Measurement

Precision measurement is the measurement that is also used to measure the performance of searching techniques and also most widely used by many researchers in order to evaluate the searching techniques adopted in their systems and different experiments. Because Precision measurement tells us that how much the number of extracted relevant records are related to the total number of relevant and irrelevant records. Following the Equation 5.2 describes the Precision measurement.

Precision = $\frac{TP}{TP + FP}$ ..........................(5.2)

5.1.4 Accuracy Score Measurement

Accuracy score measurement is a measurement that is used to check the accuracy performance of any system or experiment. It is also most widely used by many researchers in order to evaluate the accuracy performance in their systems or experiments. Accuracy value tells us that how much our systems or experiments are closed to the actual target. However, the accuracy of any system or experiment depends on upon the measuring criteria by the researcher or scientist. Because every researcher and scientists set their own actual targets for their systems in order to analyze the accuracy of the systems or experiments. There is the following Equation 5.3 that describes the Accuracy Score Measurement.

$\text{ACCURACY} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$ ..........................(5.3)

Where

TF is relevant document retrieved
TN is irrelevant document not retrieved
FP irrelevant document retrieved
FN relevant document not retrieved

5.1.5 F-Measurement

F-Measure for the proposed QA system, in f-measure two measures precision and recall are sometimes used together in the F1 Score (or f-measure) to provide a single measurement for a system.

F-Measurement for the proposed QA system that “it is double of the product of the
Recall value, Precision value to the sum of the precision and recall values for the system”. Since values of these measurements are already expressed. There is the following Equation 5.4 that describes the F-measure.

\[ F = \frac{2PR}{P + R} \] ..................(5.4)

### 5.1.6 Specificity Measurement

Sensitivity and specificity are statistical measures of the performance of a binary classification test, also known in statistics as classification function. Specificity also called the true negative rated and measured the proportion of negatives that are correctly identified as shown in Equation 5.5.

\[ \text{Specificity} = \frac{TN}{TN + FN} \] ..................(5.5)

### 5.1.7 Matthews Correlation Coefficient

The Matthews correlation coefficient is used as a measure of the quality of binary (two-class) classifications [70]. It takes into account true and false positives and negatives and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes as shown in Equation 5.6.

\[
MCC = \frac{(TP \times TN \times FP \times FN)}{((TP + FP)(TP + FN)(TN + FP)(TN + FN))} ... \ (5.6)
\]

### 5.2 Multimedia and Agent based Question Answering System Analysis

A computer program was written to send questions one by one to AnswerBus, START, LCC, IONAUT, QuASM and MAQAS; and also retrieve the answers from the systems. After the answer to a question had been received, a time interval of 10 seconds was provided before next question was sent to ensure that a performance of the system would not be affected by its work on the previous one. The response times to each question by the systems and the lengths of the returned answers were recorded. In order to minimize the impact of network performance on the variations of the systems response time, the answers from these systems were retrieved at the same time using a same computer. The accuracy performance for the answers to the types of questions of different QA systems. Table 5.1 describes that measure the recall values, precision values, accuracy score, F-measure, specificity, and Matthews correlation coefficient (MCC) of available online QA
systems including the proposed QA system. Basically, the compare of our proposed QA system with other QA systems for evaluating the accuracy performance of the proposed “MAQAS system”. Comparison of the accuracy performance measure of this proposed QA system with the other existing QA systems for the answers of definition, description and factoid type of questions as shown in Figure 5.1.

Table 5.1: Performance Analysis with other Question Answering Systems

<table>
<thead>
<tr>
<th>Systems Name</th>
<th>Correct TOP5</th>
<th>Correct TOP1</th>
<th>NIST Score</th>
<th>T max(s)</th>
<th>T min(s)</th>
<th>T mean(s)</th>
<th>T std dev</th>
<th>Lmean (byte)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AnswerBus</td>
<td>141</td>
<td>120</td>
<td>64.18%</td>
<td>15.06</td>
<td>3.79</td>
<td>7.2</td>
<td>3.07</td>
<td>141</td>
</tr>
<tr>
<td>IONAUT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>44.88</td>
<td>2.78</td>
<td>12.51</td>
<td>6.81</td>
<td>1312</td>
</tr>
<tr>
<td>LCC</td>
<td>97</td>
<td>75</td>
<td>41.73%</td>
<td>342.52</td>
<td>4.3</td>
<td>44.24</td>
<td>32.63</td>
<td>178</td>
</tr>
<tr>
<td>QuASM</td>
<td>13</td>
<td>7</td>
<td>4.45%</td>
<td>284.29</td>
<td>2.61</td>
<td>20.72</td>
<td>33.92</td>
<td>1766</td>
</tr>
<tr>
<td>START</td>
<td>29</td>
<td>29</td>
<td>14.50%</td>
<td>62.07</td>
<td>2.02</td>
<td>9.84</td>
<td>7.45</td>
<td>0</td>
</tr>
<tr>
<td>MAQAS</td>
<td>146</td>
<td>140</td>
<td>63.12%</td>
<td>21.5</td>
<td>1.5</td>
<td>6.19</td>
<td>3.76</td>
<td>140</td>
</tr>
</tbody>
</table>

Figure 5.1: Comparison of the QA System in terms of Response and Accuracy

5.3 Accuracy Comparison for the Image Type Questions

For the image type question performance analysis is done with Photo based Question Answering System. Numeric statistics are presented in Table 5.2. The result are shown in Figure 5.4.
### Table 5.2: MAQAS vs Photo-Based QA System Image Questions

<table>
<thead>
<tr>
<th>Types of Questions</th>
<th>No. of total questions</th>
<th>No. of total correct answers</th>
<th>Accuracy Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use Case Diagrams (MAQAS System)</td>
<td>80</td>
<td>66</td>
<td>82.5</td>
</tr>
<tr>
<td>Use Case Diagrams (Photo Based QA)</td>
<td>80</td>
<td>22</td>
<td>27.5</td>
</tr>
<tr>
<td>Books (MAQAS System)</td>
<td>80</td>
<td>71</td>
<td>88.75</td>
</tr>
<tr>
<td>Books (Photo Based QA)</td>
<td>80</td>
<td>61</td>
<td>76.25</td>
</tr>
</tbody>
</table>

### 5.4 Comparison of the QA System in terms of Response and Accuracy

This circos graphic in Figure 5.2 shows the data analysis that arranged in a circular orientation. The wedges are marked with a length scale along with different question answering system. Data placed outside of the ring represents degree of small and large scale variation in the question answer of the subset of the collect data at given position found between different populations. Data placed on top of the ring highlights, the correct answer in a sense of the answer retrieved correctly. The question answering system are discussed START, QOURA, MAQAS systems. This circos graph present the data collect real time to run the queries on these QA systems working online and their response time is measured by automatically. It provides the score in each system with response time and the correct answer or incorrect answers. The correct answers are depicted with score “1” and incorrect answers are denoted by zero.

The graph provides the numbers of correct answers in the systems top five and top one answers, the standard NIST scores, the maximal, minimal and average response times measured in seconds, standard deviation of response times, and the average lengths of returned answers as shown in Figure 5.3. The wedges are marked with a length scale along with different question answering system. Data placed outside of the ring represents the degree of small- and large- scale variation in the question answer of the subset of the collected data as shown in Table 5.1.

The maximum time, minimum time, answer top 5, answer top 1, meantime, standard deviation and mean length of the text that used to present the answer by the outside.
Figure 5.2: Accuracy Comparison of MAQAS with Online Systems

of the outer ring of the graph. The inner ring presents the mean length by the light green color, correcttop5, correcttop1 in reddish color. The time in second for maximum time and minimum time the answer in greenish yellow. The overall performance of the MAQAS system is better than the similar systems. Figure 5.4 presents the features based comparison of various question answering systems.

5.5 Analysis of MAQAS and Teacher Answers

The multimedia and agent based question answering is analyzed with answers delivered by the teacher to students. Circos software contains a complete package for information and data visualization. Data can be visualized in a circular layered layout. Circos is more ideal for exploring and investigating the associations between objects, values
Chapter 5. Results and Discussions

Figure 5.3: Comparison of the MAQAS with Photo Based QA systems

Figure 5.4: Features based Comparison of MAQAS and Existing QA Systems
or positions. Circos is also ideal for producing research article quality to present the infographics with illustrations.

Table 5.3: MAQAS Comparison with Teacher Answers

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Answering System</th>
<th>Number of Questions</th>
<th>Answer Types</th>
<th>Average Answer Length (Bytes)</th>
<th>Latency (Minutes)</th>
<th>Answer Effectiveness %</th>
<th>Learning Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Teacher</td>
<td>50</td>
<td>Factoid</td>
<td>50</td>
<td>1523</td>
<td>95</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>MAQAS</td>
<td>50</td>
<td>Factoid</td>
<td>50</td>
<td>1</td>
<td>89</td>
<td>89</td>
</tr>
<tr>
<td>3</td>
<td>Teacher</td>
<td>60</td>
<td>Definition</td>
<td>200</td>
<td>1623</td>
<td>80</td>
<td>72</td>
</tr>
<tr>
<td>4</td>
<td>MAQAS</td>
<td>60</td>
<td>Definition</td>
<td>200</td>
<td>1</td>
<td>93</td>
<td>93</td>
</tr>
<tr>
<td>5</td>
<td>Teacher</td>
<td>55</td>
<td>Descriptive</td>
<td>3000</td>
<td>1735</td>
<td>75</td>
<td>74</td>
</tr>
<tr>
<td>6</td>
<td>MAQAS</td>
<td>55</td>
<td>Descriptive</td>
<td>3000</td>
<td>2</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>7</td>
<td>Teacher</td>
<td>25</td>
<td>Image</td>
<td>400</td>
<td>1279</td>
<td>94</td>
<td>91</td>
</tr>
<tr>
<td>8</td>
<td>MAQAS</td>
<td>25</td>
<td>Image</td>
<td>400</td>
<td>3</td>
<td>89</td>
<td>89</td>
</tr>
<tr>
<td>9</td>
<td>Teacher</td>
<td>35</td>
<td>Source Code</td>
<td>1000</td>
<td>907</td>
<td>97</td>
<td>96</td>
</tr>
<tr>
<td>10</td>
<td>MAQAS</td>
<td>35</td>
<td>Source Code</td>
<td>1000</td>
<td>2</td>
<td>88</td>
<td>88</td>
</tr>
</tbody>
</table>

The fine control of each element in the figure to tailor its focus points and detail to your audience. These questions are categorized in to five prominent classes as following:

1) Factoid type questions
2) Definition type questions
3) Descriptive type questions
4) Image type questions
5) Source code type questions

The Circos graph 5.5 shows the data analysis that arranged in a circular orientation. The wedges are marked with labels along with different question answering system parameters for analysis. The parameters used during analysis are the delay factor which is effective for the learning positive or negatively. Other factors are answer effectiveness, average answer length in bytes and support for the learning. We have used the real time developed a dataset of 1108 student with 3434 questions that are answered by the teacher. We have taken 50 question of factoid type, 60 question of definition type, 55 questions descriptive type, 25 questions image type and 35 question of source code. These question and answers are analyzed carefully to find some efficient results as shown in Table 5.3.
Figure 5.5: Comparison of MAQAS with Teacher Answers

The average length of the answers is measured in bytes and latency in minutes. The out wedge represents the labels of the field data with textual information. The measure calculates for the support of analysis is learning support by the teacher answers and MAQAS system provided to the students. The learning support is calculated by using the equation as shown by the (Equation. 5.7)

\[ SL = AE - \frac{L}{AAL} \]

Where \( L \) denotes Support for learning

\( AE \) stands for Answer Effectiveness in percentage

\( L \) stands for Latency in minutes

\( AAL \) stands for Average Answer Length in bytes.
The Circos graph in Figure 5.5 presents that the performance of the teacher for the support for learning is 65 percent although answer effectiveness is 95. The inefficiency is generated due to the delay in answering the student’s question. The performance of the teacher answer is better in the case of image type questions and coding questions while the MAQAS produce better results for the factoid, descriptive and definition type questions. MAQAS accuracy and performance of learning support is better as it used the three different type of search engine. The answers generated by the Google is also handled by the system and delivered to the students.

It is more suitable for keyword extraction to be accomplished using headwords from the question. This code is written using the Python programming language. The main phrases and keywords are extracted and used to measure the overall relevance of the answer by the MAQAS. The table 5.4 presents the comparison of techniques utilized for the question answering purposes.

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Types of QA System</th>
<th>Support Multilingual</th>
<th>Single / Multiple Techniques</th>
<th>The Web Semantic Use</th>
<th>Methods Used in QA System</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A prototype QA system for answer retrieval</td>
<td>Single language</td>
<td>Single method, Measuring Semantic distance</td>
<td>Syntactic and semantic</td>
<td>Tagging and chunking, named entity recognition,</td>
</tr>
<tr>
<td>2</td>
<td>A QA System supported by information extraction</td>
<td>Single language</td>
<td>Text Processing, text matching</td>
<td>No Semantics</td>
<td>Named-entity supported QA, question processing,</td>
</tr>
<tr>
<td>3</td>
<td>Analysis of the Asks QA system</td>
<td>Single language</td>
<td>N-Gram Mining, N-Gram Filtering, N-Gram Tiling</td>
<td>Semantics</td>
<td>Question reformulation,</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>-----------------</td>
<td>-----------------------------------------------</td>
<td>-----------</td>
<td>------------------------</td>
</tr>
<tr>
<td>4</td>
<td>QA, answer type identification</td>
<td>Multilingual</td>
<td>Tokenization and POS tagging, word sense disambiguation</td>
<td>Semantics</td>
<td>Keywords expansion</td>
</tr>
<tr>
<td>5</td>
<td>A specifiable domain multilingual question</td>
<td>Multilingual</td>
<td>Single technique</td>
<td>Semantics</td>
<td>Answering architecture</td>
</tr>
<tr>
<td>6</td>
<td>Multilingualism</td>
<td>Multilingual</td>
<td>Spatial-temporal context awareness</td>
<td>Textual entailment</td>
<td>Answering architecture</td>
</tr>
<tr>
<td>7</td>
<td>A QA system based on IR and validation</td>
<td>Single Language</td>
<td>Multiple methods</td>
<td>Semantics</td>
<td>Expected answer type, named entities presence, acronym checking</td>
</tr>
<tr>
<td>8</td>
<td>A hybrid QA system based on IR</td>
<td>Single language</td>
<td>Module, document processing, and indexing</td>
<td>Answer validation</td>
<td>Pattern generation module, hypothesis generation</td>
</tr>
</tbody>
</table>
5.6 Conclusion

As MAQAS will try to minimize the time delay. The features comparison of the MQAQAS with Existing Systems is represented in Table 5.5. MAQAS optimized the solution of the student questions that answered more relevant. The proposed system will be an intelligent Question Answering System, whose search will consider the more relevance than only keywords.

<table>
<thead>
<tr>
<th>Question Answering System</th>
<th>Question Reformulation</th>
<th>Wikipedia</th>
<th>Google</th>
<th>Other Search engine</th>
<th>Answer Extraction</th>
<th>Multiple Keyword Search</th>
<th>Image Search</th>
<th>Factoid type</th>
<th>Definition type</th>
<th>Description Type</th>
<th>Open Domain</th>
<th>Answer Relevancy</th>
<th>Knowledgebase</th>
<th>Answer Ranking</th>
<th>Student Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>START QA since 1993</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>TREC QA 2006</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Question Set</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
</tr>
<tr>
<td>Answer Bus QA</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>YodaQA Live 2005</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>QOURA QA 2009</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>AskMSR QA 2002</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>MAQAS</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Chapter 6

CONCLUSION AND FUTURE WORK

This dissertation has explored one of the important problems of eLearning and proposed a solution for it. This chapter concludes the dissertation with the discussion of limitations of the proposed approach and the outline of possible directions for future research.

6.1 Conclusion

This research concluded that the automatic question answer should be integrated with eLearning Paradigm. The question answering is needed is more emphasized by the delayed assessment of questions posted by the students and replied by the instructors.

6.1.1 Next Generation Question Answering Systems

The question answering for the next generation is required to accommodate multimedia data contents. The future question answering systems need to required functionalities that integrate a variety of resources like metadata, images, audios, videos, natural language text, user added tags. The student may ask the question by using different modalities. The question answering system would be able to answer by the analysis of content in the account to semantic labeling, integrate and generate illustration for all content types, integrate user generated metadata possibilities in processing answers. The metadata may be added at content creation time lead towards cross-media content and cross lingual data with the multimedia repositories as source of information.
6.1.2 Delay Assessments

This research aims to measure and investigate the delay of the student questions. These questions are very significant for the student to continue the learning process if they are answered well on time. The learning curve is affected negatively and the delay is an annoying factor for the student. The learning curve of the online students commonly not cross the traditional learning student due to striving for answers always. This thrust is never fulfilled. The conclusion of the study is that the administrator required to hire the faculty for smooth answering the student questions. The educator and instructor should have the essential skill for answering the student. The information and communication technologies also play an important role in finding the correct and precise answer. The future directions are: to introduce techniques for the elimination of delay automatically and prioritize the questions. The machine learning approaches can also be used for the prediction of required student ratio in science education subjects. This may also be calculated for the optimal student-teacher requirement in the forthcoming enrollments.

6.1.3 Multimedia and Agent-Based Question Answering System

This MAQAS is beneficial for both instructors and students as it eliminates the delay in answering questions. Students may learn at their own pace, free from the time and location constraints. It is innovation, that offers improved and upgraded engineering-based eLearning Frameworks for instructors and students. It reduces the reaction time to students questions, and it offers a spontaneous QA session through the electronic eLearning platform, so the collaboration of students is improved in order to accelerate their studying process. The delay figure in the asynchronous eLearning paradigm can be eradicated entirely by automating the student QA framework. This QA system is implemented and performance is comparatively better with answers are more prevalent and accurate. The reason of high performance on these parameter as it uses Wikipedia, Google and Webopedia search engines for retrieval of documents. The questions using the qualitative exploration technique have the capacity to show relevancy. This research on eLearning is dependent upon extensive dissection of the principle elements and so as to enhance the QA practice. In future, this MAQAS will be enhanced to function with all search engines. This MAQAS enables the student to post questions about the diagrams related to to their subject. In future, the proposed system is intended to fulfill the additional requirement of the autonomous functionality of the agents for the image
based questions.

6.2 Future Work

The study is extended towards the blended learning framework, virtual reality and augmented reality framework for the enhancement of student understanding. In future need to accommodate virtual and augmented reality based question answering.

6.2.1 Blended Learning

With the recent developments in the educational field, the trends of learning are improved from simple learning to eLearning and then leads toward blended learning. Blended Learning (BL) is the concept that combines the advantages of e-learning with traditional methods of learning that is face-to-face interaction in classrooms. Blended learning is the learning program where more than one method of concept delivery is being used to enhance the students learning and decrease the cost of the program. Sometimes a student requires to want pictures or animations to elaborate their concepts instead of simple text reading. Blended learning paradigm also has the strong need of question answering system.

6.2.2 Virtual Reality

Virtual reality (VR) is a computer technology that uses headsets, sometimes in combination with physical spaces or multi-projected environments, to generate realistic images, sounds and other sensations that simulate a user’s physical presence in a virtual or imaginary environment. Virtual environments as it allows the user to see the real world with virtual objects that are synthesized with the reality [95, 96]. Augmented Reality concept can be reflected as the RealityVirtuality Continuum [97].

6.2.3 Augmented Reality

Augmented Reality is used in the fields of medicine, engineering, military, art, robotics, manufacturing, oil and gas, civil, commerce, construction, psychological treatments. Recently Augmented Reality can be used in the field of education to enhance users perception. AR has been used to develop students understanding in the area of science and medical [98100]. AR provides a logical and smooth connection between virtual objects and the real world, allows the user to observe and understand the actual world through virtual objects. Virtual objects used in Augmented Reality are video, sounds, 3D objects, and interactivity. The main devices for Augmented Reality are Head Mounted
Displays, input devices, computers and tracking systems. In previous research, the researcher stated that AR increases content’s understand as students can experience AR to clear their concepts and learning enhancement. Contents learned through AR are memorized more strongly as compared to non-AR experiences. The implementation of augmented reality in education and training is need of time. The question answering system integrated with augmented reality for the learning of students will also strengthening their satisfaction.
Chapter 7

REFERENCES