A PROJECT REPORT

ON

Automatic Text Categorization of Word Problems

SUBMITTED BY

BY

Mr. Suraj Kumar, B120084330

Mr. Pranav Kanade, B120084257

Mr. Gaurav Shukla, B120084237

Mr. Sanket Deshmukh, B120084228

Under the guidance of

Mrs. Shanthi K. Guru

In partial fulfillment of the requirements for

Bachelors Degree in Computer Engineering of

SAVITRIBAI PHULE PUNE UNIVERSITY

[2016-2017]

Department of Computer Engineering

D. Y. PATIL COLLEGE OF ENGINEERING

Akurdi, PUNE-411044.

D. Y. Patil College of Engineering

Akurdi, Pune-411044

Department of Computer Engineering

CERTIFICATE

This is to certify that the Project Entitled

Automatic Text Categorization of Word Problems

Submitted by

Mr. Suraj Kumar, B120084330

Mr. Pranav Kanade, B120084257

Mr. Gaurav Shukla, B120084237

Mr. Sanket Deshmukh, B120084228

is a bonafide work carried out by Students under the supervision of **Mrs. Shanthi K. Guru** and it is submitted towards the partial fulfillment of the requirement of Bachelor of Engineering (Computer Engineering).

Mrs. Shanthi K. Guru	Dr.Neeta Deshpande
Internal Guide	H.O.D
Dept. of Computer Engg.	Dept. of Computer Engg.

Principal

D. Y. Patil College of Engineering, Akurdi, Pune-44

Internal Examiner **External Examiner**

PROJECT APPROVAL SHEET

Automatic Text Categorization of Word Problems

Is successfully completed by

Mr. Suraj Kumar, B120084330

Mr. Pranav Kanade, B120084257

Mr. Gaurav Shukla, B120084237

Mr. Sanket Deshmukh, B120084228

at

DEPARTMENT OF COMPUTER ENGINEERING

D. Y. PATIL COLLEGE OF ENGINEERING, AKURDI, PUNE-44

SAVITRIBAI PHULE PUNE UNIVERSITY,PUNE

ACADEMIC YEAR 2016-2017

Mrs. Shanthi K. Guru	Dr. Neeta Deshpande
Internal Guide	H.O.D
Dept. of Computer Engg.	Dept. of Computer Engg.

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on ***'BE PROJECT TITLE'****.*

I would like to take this opportunity to thank my internal guide ***Prof. Guide Name*** *for giving me all the help and guidance I needed. I am really grateful to them for their kind support. Their valuable suggestions were very helpful.*

I am also grateful to ***Prof. HOD Name****, Head of Computer Engineering Department, CollegeName for his indispensable support, suggestions.*

In the end our special thanks to ***Other Person Name*** *for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for Our Project.*

Student Name1
Student Name2
Student Name3
Student Name4
(B.E. Computer Engg.)

ABSTRACT

Automatic Mathematical Word Problem Solver serves two purposes: First is in Intelligent Tutoring Systems and Second is in Cognitive Science to study way of thinking to solve these problems in children. To solve a MWP, First we need to classify the given problem into either Joint and Seperate problem or Part-Part-Whole problem or Compare problem. After clssification is done we need to recognize the function of each sentence and extract information from them. This extracted feature is finally used to form equations and then equations are solved to obtain the solution.Sothe proposed system consists of following main modules: Classification, Information Extraction, Equation Generator and Equation Solver.

Contents

List of Figures

List of Tables

List of Abbreviations

1 Synopsis

1.1 Project Title

Automatic Text Categorization of Word Problems

1.2 Project Option

Industry Sponsored

1.3 Internal Guide

Mrs. Shanthi K. Guru

1.4 Sponsorship and External Guide

Sponsor : Persistent Systems Ltd. Pune

Guide : Mr. Deshpande

1.5 Technical Keywords (As per ACM Keywords)

Please note ACM Keywords can be found : http://www.acm.org/about/class/ccs98-html

Example is given as

1. C. Computer Systems Organization
   (a) C.2 COMPUTER-COMMUNICATION NETWORKS
      i. C.2.4 Distributed Systems
         A. Client/server
         B. Distributed applications
         C. Distributed databases
         D. Network operating systems
         E. Distributed file systems
         F. Security and reliability issues in distributed applications

1.6 Problem Statement

Design a system which accepts an algebric word problem as an input and classifies it into pre-defined domain (e.g. Part-Part-Whole, Join-Seperate, Comparison, Equal Groups, Multiplicative-Compare etc.) using NLP (Information Extraction) and Machine Learning Techniques. Also generting equations for these problems and solving them automatically.

1.7 Abstract

Automatic Mathematical Word Problem Solver serves two purposes: First is in Intelligent Tutoring Systems and Second is in Cognitive Science to study way of thinking to solve these problems in children. To solve a MWP, First we need to classify the given problem into either Joint and Seperate problem or Part-Part-Whole problem or Compare problem. After clssification is done we need to recognize the function of each sentence and extract information from them. This extracted feature is finally used to form equations and then equations are solved to obtain the solution.Sothe proposed system consists of following main modules: Classification, Information Extraction, Equation Generator and Equation Solver.

1.8 Goals and Objectives

1. Building assistant tool for students to solve algebric word problems
2. To study thought process while solving a algebric word problem
3. To assist in teaching as Intelligent Tutoring System

1.9 Relevant mathematics associated with the Project

System Description:

- Input:
- Output:

- Identify data structures, classes, divide and conquer strategies to exploit distributed/parallel/concurrent processing, constraints.
- Functions : Identify Objects, Morphisms, Overloading in functions, Functional relations
- Mathematical formulation if possible
- Success Conditions:
- Failure Conditions:

1.10 Names of Conferences / Journals where papers can be published

- National Conference on Emerging Trends in Computer Engineering and Technology, 24-25 March 2017, at MIT AOE, Pune

1.11 Review of Conference/Journal Papers supporting Project idea

Atleast 10 papers + White papers or web references
Brief literature survey [Description containing important description of at least 10 papers

1.12 Plan of Project Execution

Using planner or alike project management tool.

2 Technical Keywords

2.1 Area of Project

Project Area

2.2 Technical Keywords

Please note ACM Keywords can be found : http://www.acm.org/about/class/ccs98-html

Example is given as

1. C. Computer Systems Organization
   (a) C.2 COMPUTER-COMMUNICATION NETWORKS
      i. C.2.4 Distributed Systems
         A. Client/server
         B. Distributed applications
         C. Distributed databases
         D. Network operating systems
         E. Distributed file systems
         F. Security and reliability issues in distributed applications

3 Introduction

3.1 Project Idea

- Project Idea

3.2 Motivation of the Project

- Motivation of the Project

3.3 Literature Survey

- Review of the papers, Description , Mathematical Terms

3.3.1 Discussion of base paper

3.3.2 Related Work

4 Problem Definition and scope

4.1 Problem Statement

Description of Problem

4.1.1 Goals and objectives

Goal and Objectives:

- Overall goals and objectives of software, input and output description with necessary syntax, format etc are described

4.1.2 Statement of scope

- A description of the software with Size of input, bounds on input, input validation, input dependency, i/o state diagram, Major inputs, and outputs are described without regard to implementation detail.
- The scope identifies what the product is and is not, what it will and won't do, what it will and wont contain.

4.2 Major Constraints

- Any constraints that will impact the manner in which the software is to be specified, designed, implemented or tested are noted here.

4.3 Methodologies of Problem solving and efficiency issues

- The single problem can be solved by different solutions. This considers the performance parameters for each approach. Thus considers the efficiency issues.

4.4 Outcome

- Outcome of the project

4.5 Applications

- Applications of Project

4.6 Hardware Resources Required

Sr. No.	Parameter	Minimum Requirement	Justification
1	CPU Speed	2 GHz	Remark Required
2	RAM	3 GB	Remark Required

Table 1: Hardware Requirements

4.7 Software Resources Required

Platform :

1. Operating System:
2. IDE:
3. Programming Language

5 Project Plan

6.1 Project Estimates

Use Waterfall model and associated streams derived from assignments 1,2, 3, 4 and 5(Annex A and B) for estimation.

6.1.1 Reconciled Estimates

6.1.2 Cost Estimate

6.1.3 Time Estimates

6.2 Risk Management w.r.t. NP Hard analysis

This section discusses Project risks and the approach to managing them.

6.2.1 Risk Identification

For risks identification, review of scope document, requirements specifications and schedule is done. Answers to questionnaire revealed some risks. Each risk is categorized as per the categories mentioned in [**?**]. Please refer table 2 for all the risks. You can refereed following risk identification questionnaire.

1. Have top software and customer managers formally committed to support the project?
2. Are end-users enthusiastically committed to the project and the system/product to be built?
3. Are requirements fully understood by the software engineering team and its customers?
4. Have customers been involved fully in the definition of requirements?
5. Do end-users have realistic expectations?
6. Does the software engineering team have the right mix of skills?

7. Are project requirements stable?
8. Is the number of people on the project team adequate to do the job?
9. Do all customer/user constituencies agree on the importance of the project and on the requirements for the system/product to be built?

6.2.2 Risk Analysis

The risks for the Project can be analyzed within the constraints of time and quality

ID	Risk Description	Probability	Impact		
			Schedule	Quality	Overall
1	Description 1	Low	Low	High	High
2	Description 2	Low	Low	High	High

Table 2: Risk Table

Probability	Value	Description
High	Probability of occurrence is	$> 75\%$
Medium	Probability of occurrence is	$26 - 75\%$
Low	Probability of occurrence is	$< 25\%$

Table 3: Risk Probability definitions [**?**]

6.2.3 Overview of Risk Mitigation, Monitoring, Management

Following are the details for each risk.

Impact	Value	Description
Very high	$> 10\%$	Schedule impact or Unacceptable quality
High	$5 - 10\%$	Schedule impact or Some parts of the project have low quality
Medium	$< 5\%$	Schedule impact or Barely noticeable degradation in quality Low Impact on schedule or Quality can be incorporated

Table 4: Risk Impact definitions [**?**]

Risk ID	1
Risk Description	Description 1
Category	Development Environment.
Source	Software requirement Specification document.
Probability	Low
Impact	High
Response	Mitigate
Strategy	Strategy
Risk Status	Occurred

Risk ID	2
Risk Description	Description 2
Category	Requirements
Source	Software Design Specification documentation review.
Probability	Low
Impact	High
Response	Mitigate
Strategy	Better testing will resolve this issue.
Risk Status	Identified

Risk ID	3
Risk Description	Description 3
Category	Technology
Source	This was identified during early development and testing.
Probability	Low
Impact	Very High
Response	Accept
Strategy	Example Running Service Registry behind proxy balancer
Risk Status	Identified

6.3 Project Schedule

6.3.1 Project task set

Major Tasks in the Project stages are:

- Task 1:
- Task 2:
- Task 3:
- Task 4:
- Task 5:

6.3.2 Task network

Project tasks and their dependencies are noted in this diagrammatic form.

6.3.3 Timeline Chart

A project timeline chart is presented. This may include a time line for the entire project. Above points should be covered in Project Planner as Annex C and you can mention here Please refer Annex C for the planner

6.4 Team Organization

The manner in which staff is organized and the mechanisms for reporting are noted.

6.4.1 Team structure

The team structure for the project is identified. Roles are defined.

6.4.2 Management reporting and communication

Mechanisms for progress reporting and inter/intra team communication are identified as per assessment sheet and lab time table.

7 Software requirement specification

7.1 Introduction

7.1.1 Purpose and Scope of Document

The purpose of SRS and what it covers is to be stated

7.1.2 Overview of responsibilities of Developer

What all activities carried out by developer?

7.2 Usage Scenario

This section provides various usage scenarios for the system to be developed.

7.2.1 User profiles

The profiles of all user categories are described here.(Actors and their Description)

7.2.2 Use-cases

All use-cases for the software are presented. Description of all main Use cases using use case template is to be provided.

Sr No.	Use Case	Description	Actors	Assumptions
1	Use Case 1	Description	Actors	Assumption

Table 5: Use Cases

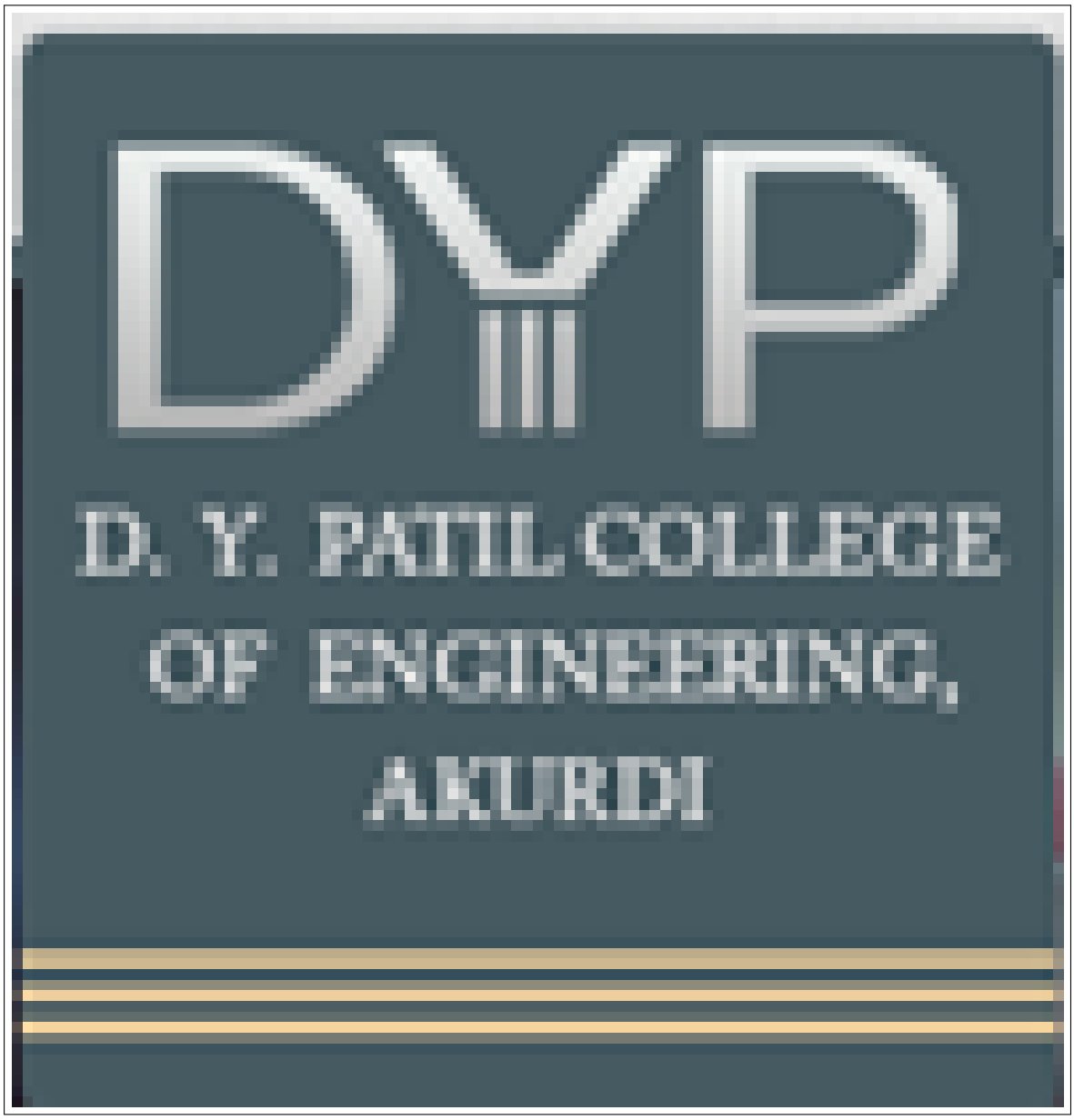

Figure 1: Use case diagram

7.2.3 Use Case View

Use Case Diagram. Example is given below

7.3 Data Model and Description

7.3.1 Data Description

Data objects that will be managed/manipulated by the software are described in this section. The database entities or files or data structures required to be described. For data objects details can be given as below

7.3.2 Data objects and Relationships

Data objects and their major attributes and relationships among data objects are described using an ERD- like form.

7.4 Functional Models and Description

A description of each major software function, along with data flow (structured analysis) or class hierarchy (Analysis Class diagram with class description for object oriented system) is presented.

7.4.1 Non Functional Requirements:

- Interface Requirements
- Performance Requirements
- Software quality attributes such as availability [related to Reliability], modifiability [includes portability, reusability, scalability] , performance, security, testability and usability[includes self adaptability and user adaptability]

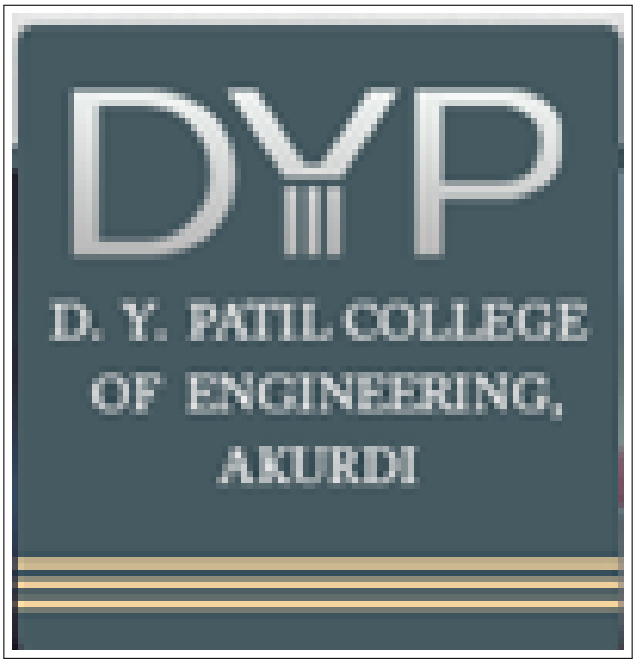

Figure 2: State transition diagram

7.4.2 Software Interface Description

The software interface(s)to the outside world is(are) described. The requirements for interfaces to other devices/systems/networks/human are stated.

8 Detailed Design Document using Appendix A and B

8.1 Introduction

8.2 UML Diagrams

This document specifies the design that is used to solve the problem of Product.

8.3 Design Constraints

Any design constraints that will impact the subsystem are noted.

8.4 Architectural Design

A description of the program architecture is presented. Subsystem design or Block diagram,Package Diagram,Deployment diagram with description is to be presented.

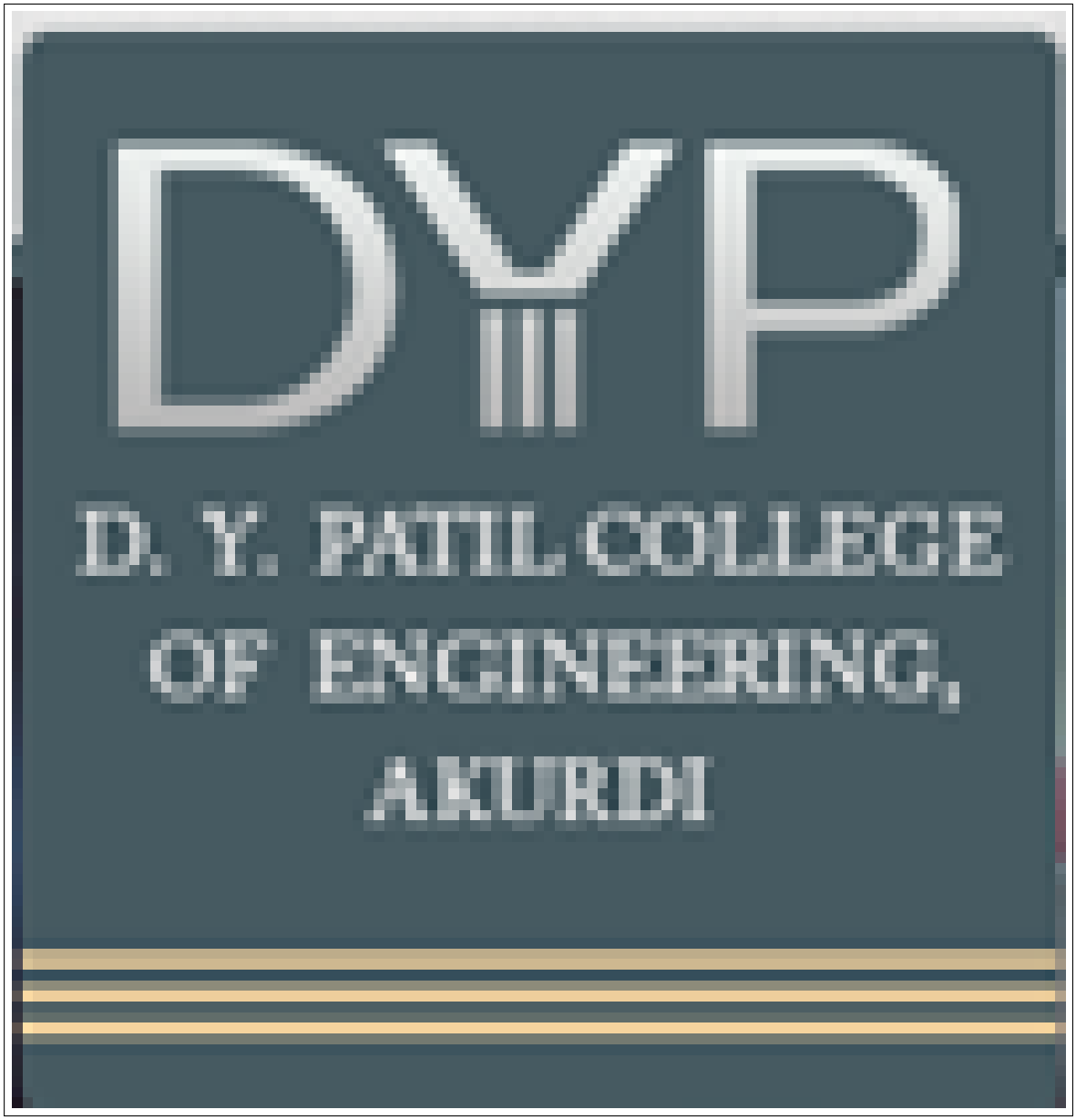

Figure 3: Architecture diagram

9 Project Implementation

9.1 Introduction

9.2 Tools,Technologies and Dataset Used

9.3 Methodologies/Algorithm Details

9.3.1 Algorithm 1/Pseudo Code

9.3.2 Algorithm 2/Pseudo Code

9.3.3 Time And Space Complexities

9.4 Verification and Validation for Acceptance

10 Software Testing

10.0.1 Testing Tools

discuss Identified and used testing tools

10.0.2 Type of Testing Used

Unit,Integration,system etc.

10.0.3 Test Cases and Test Results

for each type of testing done.

11 Results

11.1 Result table

11.2 Screen shots

Outputs / Snap shots of the results

11.3 Outputs/Graph/Charts

Outputs / Snap shots of the results

12 Deployment and Maintenance

12.1 Operational Requirements to run the System

12.2 User help

13 Conclusion and Future Scope

Write summary , conclusion and future scope

Appendix A References

(Strictly in ACM Format)

Appendix B Laboratory assignments on Project Analysis of Algorithmic Design

- To develop the problem under consideration and justify feasibilty using concepts of knowledge canvas and IDEA Matrix.

Refer [**?**] for IDEA Matrix and Knowledge canvas model. Case studies are given in this book. IDEA Matrix is represented in the following form. Knowledge canvas represents about identification of opportunity for product. Feasibility is represented w.r.t. business perspective.

I	D	E	A
Increase	Drive	Educate	Accelerate
Improve	Deliver	Evaluate	Associate
Ignore	Decrease	Eliminate	Avoid

Table 6: IDEA Matrix

- Project problem statement feasibility assessment using NP-Hard, NP-Complete or satisfy ability issues using modern algebra and/or relevant mathematical models.
- input x,output y, y=f(x)

Appendix C Laboratory assignments on Project Quality and Reliability Testing of Project Design

It should include assignments such as

- Use of divide and conquer strategies to exploit distributed/parallel/concurrent processing of the above to identify object, morphisms, overloading in functions (if any), and functional relations and any other dependencies (as per requirements). It can include Venn diagram, state diagram, function relations, i/o relations; use this to derive objects, morphism, overloading
- Use of above to draw functional dependency graphs and relevant

Software modeling methods, techniques including UML diagrams or other necessities using appropriate tools.

- Testing of project problem statement using generated test data (using mathematical models, GUI, Function testing principles, if any) selection and appropriate use of testing tools, testing of UML diagram's reliability. Write also test cases [Black box testing] for each identified functions. You can use Mathematica or equivalent open source tool for generating test data.
- Additional assignments by the guide. If project type as Entreprenaur, Refer [**?**],[**?**],[**?**], [**?**]

Appendix D Project Planner

Using planner or alike project management tool.

Appendix E Reviewers Comments of Paper Submitted

(At-least one technical paper must be submitted in Term-I on the project design in the conferences/workshops in IITs, Central Universities or UoP Conferences or equivalent International Conferences Sponsored by IEEE/ACM)

1. Paper Title:
2. Name of the Conference/Journal where paper submitted :
3. Paper accepted/rejected :
4. Review comments by reviewer :
5. Corrective actions if any :

Appendix F Plagiarism Report

Plagiarism report

Appendix G Term-II Project Laboratory Assignments

1. Review of design and necessary corrective actions taking into consideration the feedback report of Term I assessment, and other competitions/conferences participated like IIT, Central Universities, University Conferences or equivalent centers of excellence etc.
2. Project workstation selection, installations along with setup and installation report preparations.
3. Programming of the project functions, interfaces and GUI (if any) as per 1 st Term term-work submission using corrective actions recommended in Term-I assessment of Term-work.
4. Test tool selection and testing of various test cases for the project performed and generate various testing result charts, graphs etc. including reliability testing.

   Additional assignments for the Entrepreneurship Project:
5. Installations and Reliability Testing Reports at the client end.

Appendix H Information of Project Group Members

one page for each student .

1. Name :
2. Date of Birth :
3. Gender :
4. Permanent Address :
5. E-Mail :
6. Mobile/Contact No. :
7. Placement Details :
8. Paper Published :