SAVITRIBAI PHULE PUNE UNIVERSITY

A PRELIMINARY PROJECT REPORT ON

BE PROJECT TITLE

SUBMITTED TOWARDS THE
PARTIAL FULFILLMENT OF THE REQUIREMENTS OF

BACHELOR OF ENGINEERING (Computer Engineering)

BY

Ayush Gandhi	Exam No:
Ankit Bhagat	Exam No:
Manav Bharambe	Exam No:

Under The Guidance of

Prof. Y. B. Desai

Prof. P. S. Vidap

College Logo Here

DEPARTMENT OF COMPUTER ENGINEERING
Pune Institute of Computer Technology
Sr. No 27, Pune-Satara Road, Dhankawadi, Pune-411 043,
Maharashtra (India).

College Logo Here

Pune Institute of Computer Technology
DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

This is to certify that the Project Entitled

Tag Scanner-cum-price comparison for fashion products

Submitted by

Ayush Gandhi	Exam No:
Ankit Bhagat	Exam No:
Manav Bharambe	Exam No:

is a bonafide work carried out by Students under the supervision of Prof. Guide Name and it is submitted towards the partial fulfillment of the requirement of Bachelor of Engineering (Computer Engineering) Project.

Prof. Y. B. Desai	Prof. P. S. Vidap	Prof. G. P. Potdar
Internal Guide	Internal Guide	H.O.D
Dept. of Computer Engg.	Dept. of Computer Engg.	Dept. of Computer Engg.

Abstract

More options often lead to more confusion. Finding the cheapest exact product while comparing both the current shop and all the e-commerce market places is the major motive. Due to fierce competition between various e-commerce websites , in a rat race for establishing huge consumer community there is price difference to grab consumer attention. We aim to create a tool that will empower consumer to get his/her product at lowest price.

In this noisy E-commerce world , often consumer is has an opportunity to get products at lowest price but consumer is left unheard of them. Though we have the power of internet and social media still majority of common masses are left clueless of such deals . We are enabling the users to stay updated about the e-commerce world. So , we are aiming to target fashion products as that are the one's which are not easy to be searched at present. We are marching towards creating a good interface for consumers to explore product on the web easily. We are aiming to solve this general problem that was not properly handled by present e-commerce giants.

We are attempting to solve the problem by taking snapshot of the price tag.We will be achieving it by using Optical character Recognition and barcode Scanner. This will provide us with rich data which will help us in searching items over the web. Searching will be done using API's of some e-commerce market place and mining from others.

Acknowledgments

It gives us great pleasure in presenting the preliminary project report on **'Tag Scanner-cum-price comparison for fashion products'***.*

I would like to take this opportunity to thank my internal guides **Prof. Y. B. Desai and Prof. P. S. Vidap** *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. G. P. Potdar**, *Head of Computer Engineering Department, Pune Institute of Computer Technology for his indispensable support, suggestions.*

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

<div align="right">

Ayush Gandhi
Ankit Bhagat
Manav Bharambe
(B.E. Computer Engg.)

</div>

INDEX

List of Figures

List of Tables

CHAPTER 1

SYNOPSIS

1.1   PROJECT TITLE

Tag Scanner-cum-price comparison for fashion products.

1.2   PROJECT OPTION

Industry Sponsored.

1.3   INTERNAL GUIDE

Prof. Y. B. Desai and Prof. P. S. Vidap .

1.4   SPONSORSHIP AND EXTERNAL GUIDE

Sponsored By Horicent Systems .

1.5   TECHNICAL KEYWORDS (AS PER ACM KEYWORDS)

1. H. Information Systems

    (a) H.3 Information Storage And Retrieval

        i. H.3.1 Content Analysis and Indexing

            A. Dictionaries

            B. Indexing methods

            C. Linguistic processing

            D. Thesauruses

        ii. H.3.3 Information Search and Retrieval

            A. Information filtering

            B. Query formulation

            C. Search process

            D. Selection process

            E. Retrieval Models

        iii. H.3.5 Online Information Services

A. Commercial services

B. Data sharing

C. Web-based services

1.6 PROBLEM STATEMENT

We want to empower the entire consumer base with a tool that will help them to get any product they are seeking at lowest possible price. In a way we aim to help them to get best-of-all worlds. We aim to create a One-Stop Shop for price comparison.

Today we have too many e-commerce websites booming around and in a rat race of trying to create a large consumer base this Business Magnates have a single Mantra , "Flash Sales" . So , many times consumer is unaware of the fact that same product that they want to buy is available at much lower cost at other website , which will result into user ending up buying the product at much higher price.

As majority consumer base has a cell-phone we will be using Optical Character Recognition and Machine Learning to gather the data and display it.

1.7 ABSTRACT

More options often lead to more confusion. Finding the cheapest exact product while comparing both the current shop and all the e-commerce market places is the major motive. Due to fierce competition between various e-commerce websites , in a rat race for establishing huge consumer community there is price difference to grab consumer attention. We aim to create a tool that will empower consumer to get his/her product at lowest price.

In this noisy E-commerce world , often consumer is has an opportunity to get products at lowest price but consumer is left unheard of them. Though we have the power of internet and social media still majority of common masses are left clueless of such deals . We are enabling the users to stay updated about the e-commerce

world. So , we are aiming to target fashion products as that are the one's which are not easy to be searched at present. We are marching towards creating a good interface for consumers to explore product on the web easily. We are aiming to solve this general problem that was not properly handled by present e-commerce giants.

We are attempting to solve the problem by taking snapshot of the price tag.We will be achieving it by using Optical character Recognition and barcode Scanner. This will provide us with rich data which will help us in searching items over the web. Searching will be done using API's of some e-commerce market place and mining from others.

1.8 GOALS AND OBJECTIVES

- Recognize the text written on price tags describing various details about the product using optical character recognition. A phone camera will be used to capture the image of the price tag and OCR will be used to get text from this image. The price tag is assumed to have a white background and has a black text on it.

- Scan the barcode if available on the price tag to optimize the search results. The details contained within the barcode will act as input for the product searching algorithm.

- Search for the product on different e-commerce websites and return the price at which the product can be obtained from the said website.

1.9 RELEVANT MATHEMATICS ASSOCIATED WITH THE PROJECT

Let S be the system solution perspective of the given problem statement where S is,

$S = \{ \ S_t \ , e \ , X \ , Y \ , F_{me} \ , DD \ , NDD \ , S_c \ , F_c \ , / \Phi_s \ \}$

where ,

- $S_t \longrightarrow$ Start State.

  $S_t = \{ \ C \ , I_{internet} \ \}$

where ,

$C \longrightarrow$ Camera.

$C = \{\ 0\ ,1\ \}$

$1 \longrightarrow$ Camera Switched On and Ready for taking the Image.

$0 \longrightarrow$ Camera Switched Off.

$I_{Internet} \longrightarrow$ Internet Connection Required.

- $e \longrightarrow$ End State.

  $e = \{\ r_i /\ 1 \leq i \leq n,\ where\ n \in I\ \}$

  where ,

  $r_i \longrightarrow$ Represents the rates in ascending order from *Low* to *High*.

- $X \longrightarrow$ Set of Inputs.

  $X = \{\ X_1\ ,X_2\ ,X_3\ ,X_4\ \}$

  where ,

  $X_1 \longrightarrow$ Image Of Price Tag.

  $X_1 \in \{^*.jpeg\ ,^*.png\ \}$

  $X_2 \longrightarrow$ Barcode present on the price tag.

  $X_3 \longrightarrow$ Language file for OCR Engine.

  $X_3 = \{\ English\ \}$

  $X_4 \longrightarrow$ Product's information to be searched on merchant website.

  $X_4 = \{\ Tag\_Information\_Title\ :\ Tag\_Information\ \}$

  where,

  $Tag\_Information\_Title = \{\ (\ [a-z][A-Z][0-9]\ )^*\ \}$

  $Tag\_Information = \{\ (\ [a-z][A-Z][0-9]\ )^*\ \}$

- $Y \longrightarrow$ Set of Outputs.

  $Y = \{\ Y_1\ ,Y_2\ ,Y_3\ ,Y_4\ ,Y_5\ \}$

  where ,

  $Y_1 \longrightarrow$ Segmented Image.

$Y_1 = \{ y_{11}, y_{12}, y_{13}, \dots \}$

where,

$y_{11}, y_{12}, y_{13}, \dots \longrightarrow$ Segmented parts of the Image.

$Y_2 \longrightarrow$ Text obtained from image.

$Y_2 = \{ y_{21}, y_{22}, y_{23}, y_{24}, y_{25} \}$

where,

$y_{21} =$ Brand Name.

$y_{21} = \{ ([a-z][A-Z][0-9])^* \}$

$y_{22} \longrightarrow$ Price.

$y_{22} = \{ (y_{221}) (y_{222}) \}$

where ,

$y_{221} = \{ \$, \varepsilon , Rs. \}$

$y_{222} = \{ [0-9]^* \}$

$y_{23} \longrightarrow$ Size.

$y_{23} = \{ [S,M,L,XL,XXL,XXXL] \,||\, [0-9]^* \}$

$y_{24} \longrightarrow$ Color.

$y_{25} \longrightarrow$ Product.

$y_{25} \; \varepsilon \; \{$ Shirt, Trouser, Shoe, Accessory,...$\}$

$Y_3 \longrightarrow$ Prices returned from merchant websites.

$Y_3 = \{ y_{31}, y_{32}, y_{33}, \dots \}$

where,

$y_{31}, y_{32}, y_{33}, \dots \longrightarrow$ Product details from different e-Commerce giants.

$Y_4 \longrightarrow$ Recommendation of Products for a particular customer.

$Y_4 = \{ y_{41}, y_{42}, y_{43}, y_{44}, y_{45}, \dots \}$

$y_{41}, y_{42}, y_{43}, y_{44}, y_{45}, \dots \longrightarrow$ Product names and details.

- $F_{me} \longrightarrow$ Set of Functions.

  $F_{me} = \{ F_1, F_2, F_3, F_4, F_5, , F_6, F_7 \}$

  where ,

  $F_1 \longrightarrow$ This function is used to detect segments of objects on to the image.

  $F_1 = seg\_image().$

seg $_-$ image() : $X_1 \longrightarrow Y_1$.
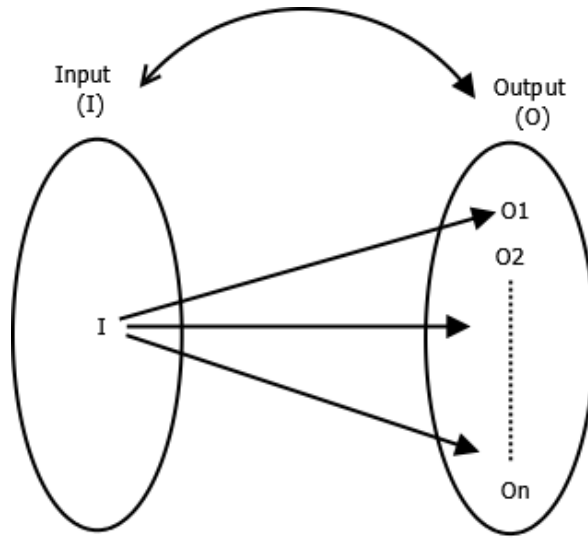
Figure 1.1: Functional Mapping

$F_2 \longrightarrow$ This function is used to convert segmented image into text.

F_2 = OCR().

OCR() : $Y_1 \longrightarrow Y_2$.

$F_3 \longrightarrow$ This function is used to scan 1-D and 2-D barcodes.

F_3 = Barcode $_-$ scanner().

Barcode $_-$ scanner() : $X_1 \longrightarrow$ 11 or 13 Digits.

$F_4 \longrightarrow$ This function is used to perform search from the text or number obtained from OCR or Barcode scanner respectively.

F_4 = search $_-$ apis().

search $_-$ apis() : $Y_2 \longrightarrow Y_3$ & $Y_4 \; || \; (J \, || \, X)$.

where ,

$J \longrightarrow$ Data in jason format.

$X \longrightarrow$ Data in XML format.

$F_5 \longrightarrow$ This function is used to recommend people a product, which would most likely be their next purchase.

F_5 = recommend $_-$ product().

recommend _ product() : $PL \longrightarrow \{LORP\}$.

where ,

PL $\longrightarrow$ Product list.

LORP $\longrightarrow$ List of Recommended Products.
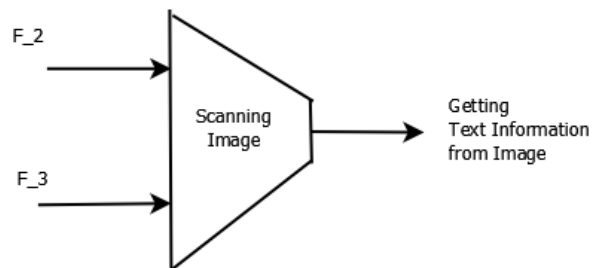
Figure 1.2: Functions that can be Morphed
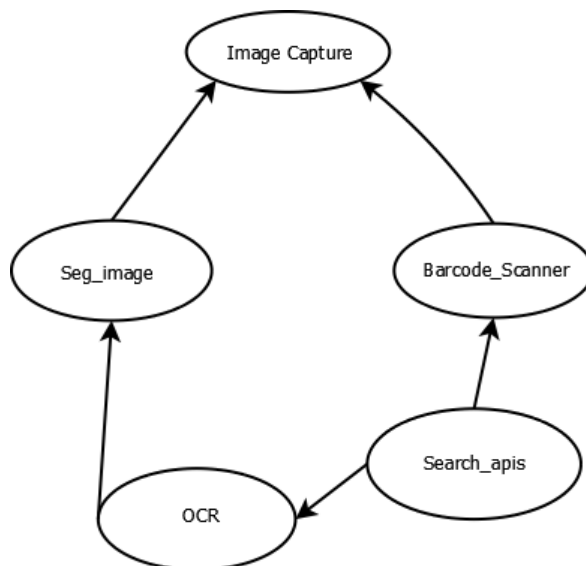
Figure 1.3: Functional Dependency Diagram

- DD $\longrightarrow$ Deterministic data.

  DD = $\{ Y_2 \}$

  ▷ Data from barcode scanner.

- NDD $\longrightarrow$ Non-Deterministic data.

  ▷ Image of price tag with background other than white.

▷ Script of text in price tag other than that of English.

- $S_c \longrightarrow$ Success Case.

  $S_c = \{ \, Y \, \}$

  ▷ Word not guessed correctly via OCR.

  ▷ Merchant website down.

  ▷ Product not available on a merchant website.

- $F_c \longrightarrow$ Failure Case.

  ▷ Successful image to text conversion through OCR.

  ▷ Proper and appropriate products searched for on merchant website.

  ▷ Successful barcode scanning done.

- System Description

Q0 ---> Start State
Q1 ----> Barcode Scanner
Q2 ----> Image Segmentation State
Q3 ----> Optical Character Recognition
Q4 ----> Searching State
Q5 ----> Display State
Q6 ----> End State
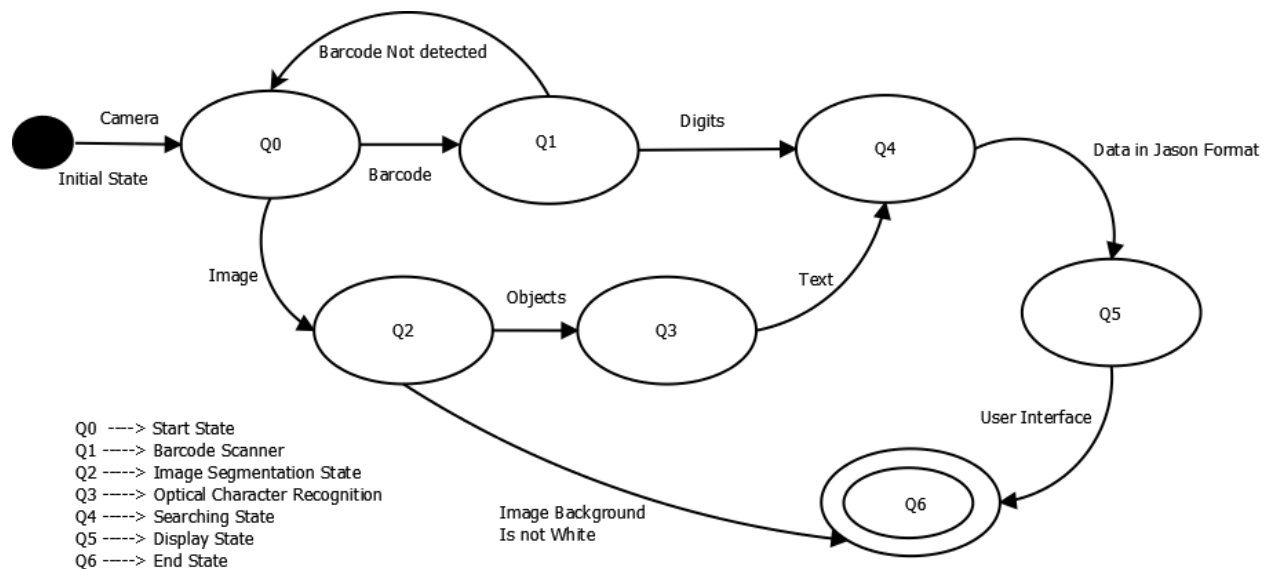
Figure 1.4: State Diagram

1.10 NAMES OF CONFERENCES / JOURNALS WHERE PAPERS CAN BE PUBLISHED

- **ICML** - International Conference on Machine Learning

- Conferences/workshops in IITs

- **ICIP** - International Conference on Image Processing

- **KDD** - Knowledge Discovery and Data Mining

1.11   REVIEW OF CONFERENCE/JOURNAL PAPERS SUPPORTING PROJECT IDEA

- **Ayatullah Faruk Mollah, Nabamita Majumder, Subhadip Basu and Mita Nasipuri, Design of an Optical Character Recognition System for Camera-based Handheld Devices, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011.**

  This paper presents a complete Optical Character Recognition (OCR) system for camera captured image/graphics embedded textual documents for handheld devices. At first, text regions are extracted and skew corrected. Then, these regions are binarized and segmented into lines and characters. Characters are passed into the recognition module.

- **Amarjot Singh, Ketan Bacchuwar, and Akshay Bhasin, A Survey of OCR Applications, Internation Journal of Machine Learning and Computing, Vol.2, No. 3, June 2012.**

  Optical Character Recognition or OCR is the electronic translation of handwritten,typewritten or printed text into machine translated images. It is widely used to recognize and search text from electronic documents or to publish the text on a website. The paper presents a survey of applications of OCR in different fields and further presents the experimentation for three important applications such as Captcha, Institutional Repository and Optical Music Character Recognition. We make use of an enhanced image segmentation algorithm based on histogram equalization using genetic algorithms for optical character recognition. The paper will act as a good literature survey for researchers starting to work in the field of optical character recognition.

- **Ravina Mithe, Supriya Indalkar, Nilam Divekar, Optical Character Recog-**

nition, International Journal of Recent Technology and Engineering(IJRTE)
The Optical Character Recognition is a mobile application. It uses smart mobile phones of android platform. This paper uses the functionality of Optical Character Recognition.

- **Yehuda Koren, Yahoo Research & Robert Bell and Chris Volinsky, Bell Labs, Matrix Factorization Techniques for Recommender Systems,Published in ACM Journal Computer, Volume 48.**
This paper recommends using Matrix Factorization models, which are superior to the nearest neighbor classic techniques for producing product recommendations,allowing the incorporation of additional information such as implicit feedback,temporal effect and confidence levels.

- **Andrew McCallum, Kamal Niggam, Jason Rennie, Kristie Seymore, A Machine Learning Approach to Domain Specific Search Engines, IJCAI99, Proceedings of the 16th international joint conference on Artificial Intelligence.**
This paper is all about creating domain specific search engines, because they offer increased accuracy and extra features not possible with the general search engine. Bayesian text Classification.

- **Seungmin Rho, Byeong-jun Han, Eenjun Hwang, SVR-based music mood classification and context based music recommendation, Proceeding MM09 Proceedings of the 17th ACM International Conference on Multimedia.**
This paper focuses on the implementation issues of context-based mood classification and music recommendation. For mood classification, they reformulate it into a regression problem based on support vector regression (SVR) and achieve 87.8 percent accuracy.

- **Web References**: Flipkart Affiliate API, (N.D)
http://www.flipkart.com/affiliate/apifaq.

- **Web References**: Amazon ItemSearch, (N.D)

http://docs.aws.amazon.com/AWSECommerceService/latest/DG/ItemSearch.html.

- **Web References**: Programmable Web API, (N.D)

  http://www.programmableweb.com/.

1.12 PLAN OF PROJECT EXECUTION

Sr. No.	Deliverables	Start Date	End Date
1	Group Formation	25/12/2014	27/12/2014
2	Horicent Sponsorship	05/07/2015	29/07/2015
3	Guide Allocation	31/07/2015	31/07/2015
4	Ideation Stage	31/07/2015	13/08/2015
5	Understanding the problem statement	13/08/2015	17/08/2015
6	Literature Survey	18/08/2015	18/09/2015

Table 1.1: Plan of Project Execution

CHAPTER 2

TECHNICAL KEYWORDS

2.1 AREA OF PROJECT

Applied Computing.

Our project mainly aims at solving a problem that is faced by masses in their day to day life that is buying things at lowest price where ever possible and try to make their buying experience awesome.

2.2 TECHNICAL KEYWORDS

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

Example is given as

1. H. Information Systems

    (a) H.3 Information Storage And Retrieval

        i. H.3.1 Content Analysis and Indexing

            A. Dictionaries

            B. Indexing methods

            C. Linguistic processing

            D. Thesauruses

        ii. H.3.3 Information Search and Retrieval

            A. Information filtering

            B. Query formulation

            C. Search process

            D. Selection process

            E. Retrieval Models

        iii. H.3.5 Online Information Services

            A. Commercial services

            B. Data sharing

            C. Web-based services

CHAPTER 3

INTRODUCTION

3.1 PROJECT IDEA

- We are empowering consumer with tool that help's him/her to get cheapest product from e-commerce giants.

3.2 MOTIVATION OF THE PROJECT

- The motivation behind this was to provide user with a nice easy interface to reduce his/her effort in getting the products at least cost possible. This will help consumer to leverage the use of opportunity provided by e-commerce websites in a race to achieve large consumer community.

3.3 LITERATURE SURVEY

- **Ayatullah Faruk Mollah, Nabamita Majumder, Subhadip Basu and Mita Nasipuri, Design of an Optical Character Recognition System for Camera-based Handheld Devices, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011.**
  This paper presents a complete Optical Character Recognition (OCR) system for camera captured image/graphics embedded textual documents for handheld devices. At first, text regions are extracted and skew corrected. Then, these regions are binarized and segmented into lines and characters. Characters are passed into the recognition module.

- **Amarjot Singh, Ketan Bacchuwar, and Akshay Bhasin, A Survey of OCR Applications, Internation Journal of Machine Learning and Computing, Vol.2, No. 3, June 2012.**
  Optical Character Recognition or OCR is the electronic translation of handwritten,typewritten or printed text into machine translated images. It is widely used to recognize and search text from electronic documents or to publish the text on a website. The paper presents a survey of applications of OCR in different fields and further presents the experimentation for three important applications such as Captcha, Institutional Repository and Optical Music

Character Recognition. We make use of an enhanced image segmentation algorithm based on histogram equalization using genetic algorithms for optical character recognition. The paper will act as a good literature survey for researchers starting to work in the field of optical character recognition.

- **Ravina Mithe, Supriya Indalkar, Nilam Divekar, Optical Character Recognition, International Journal of Recent Technology and Engineering(IJRTE)**
  The Optical Character Recognition is a mobile application. It uses smart mobile phones of android platform. This paper uses the functionality of Optical Character Recognition.

- **Yehuda Koren, Yahoo Research & Robert Bell and Chris Volinsky, Bell Labs, Matrix Factorization Techniques for Recommender Systems,Published in ACM Journal Computer, Volume 48.**
  This paper recommends using Matrix Factorization models, which are superior to the nearest neighbor classic techniques for producing product recommendations,allowing the incorporation of additional information such as implicit feedback,temporal effect and confidence levels.

- **Andrew McCallum, Kamal Niggam, Jason Rennie, Kristie Seymore, A Machine Learning Approach to Domain Specific Search Engines, IJCAI99, Proceedings of the 16th international joint conference on Artificial Intelligence.**
  This paper is all about creating domain specific search engines, because they offer increased accuracy and extra features not possible with the general search engine. Bayesian text Classification.

- **Seungmin Rho, Byeong-jun Han, Eenjun Hwang, SVR-based music mood classification and context based music recommendation, Proceeding MM09 Proceedings of the 17th ACM International Conference on Multimedia.**
  This paper focuses on the implementation issues of context-based mood classification and music recommendation. For mood classification, they reformulate it into a regression problem based on support vector regression (SVR) and achieve 87.8 percent accuracy.

- **Web References**: Flipkart Affiliate API, (N.D)

  http://www.flipkart.com/affiliate/apifaq.

- **Web References**: Amazon ItemSearch, (N.D)

  http://docs.aws.amazon.com/AWSECommerceService/latest/DG/ItemSearch.html.

- **Web References**: Programmable Web API, (N.D)

  http://www.programmableweb.com/.

CHAPTER 4

PROBLEM DEFINITION AND SCOPE

4.1 PROBLEM STATEMENT

We want to empower the entire consumer base with a tool that will help them to get any product they are seeking at lowest possible price. In a way we aim to help them to get best-of-all worlds. We aim to create a One-Stop Shop for price comparison. Today we have too many e-commerce websites booming around and in a rat race of trying to create a large consumer base this Business Magnates have a single Mantra , Flash Sales . So , many times consumer is unaware of the fact that same product that they want to buy is available at much lower cost at other website, which will result into user ending up buying the product at much higher price. As majority consumer base has a cell-phone we will be using Optical Character Recognition and Machine Learning to gather the data and display it.

4.1.1 Goals and objectives

Goal and Objectives:

- Recognize the text written on price tags describing various details about the product using optical character recognition. A phone camera will be used to capture the image of the price tag and OCR will be used to get text from this image. The price tag is assumed to have a white background and has a black text on it.

- Scan the barcode if available on the price tag to optimize the search results. The details contained within the barcode will act as input for the product searching algorithm.

- Search for the product on different e-commerce websites and return the price at which the product can be obtained from the said website.

4.1.2 Statement of scope

- Image of the price tag is the input. This image should have a white background with black text written on it. The image should be clear so that the OCR can successfully recognize the characters in the given image. The output of the

OCR is the various words contained in the image of the price tag. This acts as an input for the search algorithm.

- The tag scanner can compare the rates of the given product at the current store with various online options. It cannot compare the rates at different physical stores. It is considered to be the future scope where we can compare the prices of the product even at different offline stores.

Q0 ---> Start State
Q1 ----> Barcode Scanner
Q2 ----> Image Segmentation State
Q3 ----> Optical Character Recognition
Q4 ----> Searching State
Q5 ----> Display State
Q6 ----> End State
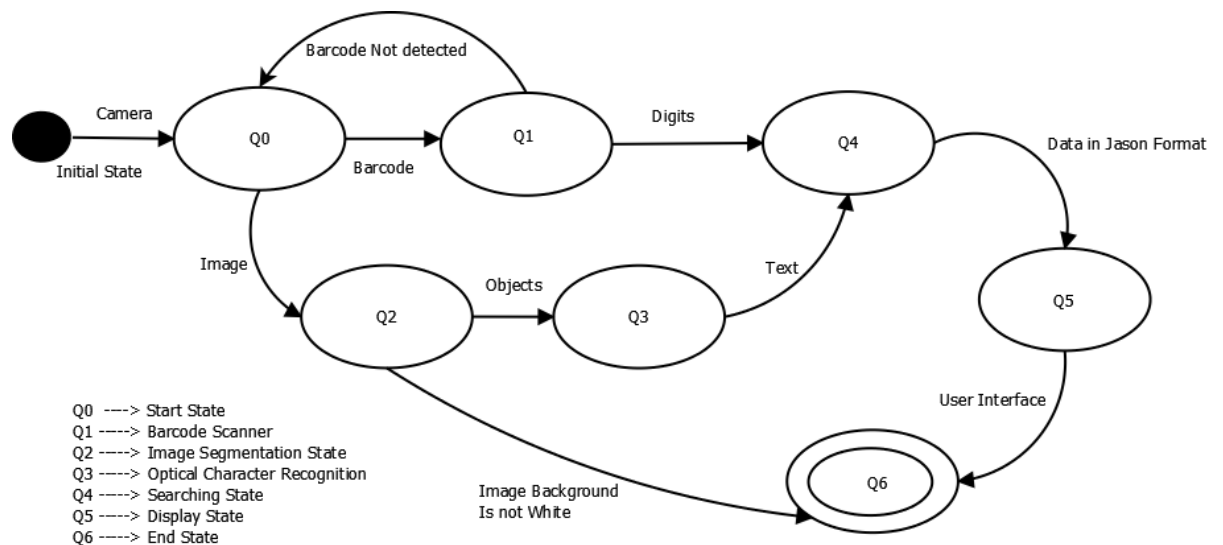
Figure 4.1: I/O State Diagram

4.2 SOFTWARE CONTEXT

- The e-world's lowest price finder.

- On the go product finder at your finger tips using just the utilities of your smart phone.

4.3 MAJOR CONSTRAINTS

- The image captured of the price tag is the bottle neck of this software. If the image is clear, OCR return accurate results. But if the image captured is blurred or unclear, OCR will return inefficient results resulting in wrong input to the search algorithm.

Table 4.1: Percentage of recognition with different classifiers

Classifier	No. of Test Patterns	No. of Correct Classification	% of Recognition
Template Matching (62 class)	15807	14659	92.74%
Template Matching (73 class)	15807	13089	82.81%
Tesseract	15807	14781	93.51%

- Availability of different websites as, if the website is down, price result cannot be obtained.

4.4 METHODOLOGIES OF PROBLEM SOLVING AND EFFICIENCY ISSUES

- Image to text conversion (Optical character Recognition) can be done by various ways. A comparison of different classifiers are given below.

- When the exact product is unavailable, we plan to implement a recommendor system. Instead of using nearest neighbor techniques, matrix factorization will be used. Matrix factorization models are superior to classic nearest-neighbor techniques for producing product recommendations, allowing the incorporation of additional information such as implicit feedback, temporal effects, and confidence levels.

4.5 OUTCOME

- Text to image conversion

- Data mining of product details on different websites.

- Recommend products when the exact product is unavailable.

4.6   APPLICATIONS

- Find minimum price of the product which was liked by user at a physical shop.

4.7   HARDWARE RESOURCES REQUIRED

Sr. No.	Parameter	Minimum Requirement	Justification
1	CPU Speed	2 GHz	The various softwares require at least this much CPU speed.
2	RAM	3 GB	Android Studio requires minimum 2 GB RAM and 4 GB RAM is recommended.
3	Android Phone	5MP Camera	Phone is required for testing and for image clarity.

Table 4.2: Hardware Requirements

4.8   SOFTWARE RESOURCES REQUIRED

Platform :

1. Operating System: 64 bit Fedora 21

2. Android Studio

3. Java, OpenCV, Python

CHAPTER 5

PROJECT PLAN

5.1 PROJECT ESTIMATES

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

5.1.1 Reconciled Estimates

5.1.1.1 Cost Estimate

5.1.1.2 Time Estimates

5.1.2 Project Resources

Project resources [People, Hardware, Software, Tools and other resources] based on Memory Sharing, IPC, and Concurrency derived using appendices to be referred. Various types of project resources are being utilised, which can be majorly categorised into the following types:

- People → Various people are involved in making this project, being specialised in different fields, overall helping to make it a complete product.

  1. Ankit Bhagat

  2. Ayush Gandhi

  3. Manav Bharambe

  4. Akash Agrawal

  5. Vrushali

- Hardware and Software → Hardware involves the phones which would be majorly used at the end for interacting with the potential customers or people on a general basis. About software, various tools and APIs would be used to support the product.

  - Hardware −

    Laptop/Desktop Machines with atleast

    a) RAM: 2GB

    b) HDD : 500 GB

c) Processor: Intel Pentium

d) Android Phone

e) iPhone

– Software −

a) LINUX based Operating System

b) Mac OS X

c) Editors like Sublime, gedit, nano.

- Tools → Tools include various APIs, image processing tools, Application development tools and deployment tools.

    – OCR for Image Processing

    – Amazon ItemSearch API and FlipkartSearch API

    – Android Studio / Eclipse

    – Xcode 7

5.2 RISK MANAGEMENT W.R.T. NP HARD ANALYSIS

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

5.2.1 Risk Identification

Identifying risks is one of the major processes involved in making a certain project. The questionnaire helped to identify them and would remind us to tackle them.

1. The user interface should be really user friendly so that the end users stick to the product and the project continues to flourish.

2. The great improvement in the hardware of the phones being used nowadays have reduced the risk of clarity in images to a certain extent, but OCR being a still risk at identification.

3. End users finding a cheaper product at a different place would basically lead to a negative review of the product. So producing a result which is favourable for the customer should be the main concern always.

4. The number of people working on this project are adequate, one more added on the group would help to make the product faster.

5.2.2 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	OCR Detection	Medium	Low	High	High
2	Website Changed Structure	Low	Low	High	High

<div align="center">Table 5.1: Risk Table</div>

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

<div align="center">Table 5.2: Risk Probability definitions [?]</div>

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

<div align="center">Table 5.3: Risk Impact definitions [?]</div>

5.2.3 Overview of Risk Mitigation, Monitoring, Management

Following are the details for each risk.

Risk ID	1
Risk Description	OCR Identification
Category	Development tools.
Source	Identified during early testing.
Probability	Medium
Impact	High
Response	Mitigate
Strategy	Combination of Segmentation, OCR and Word Predict.
Risk Status	Occurred

Risk ID	2
Risk Description	Support for devices
Category	Requirements
Source	Software Design Specification documentation review.
Probability	Low
Impact	High
Response	Mitigate
Strategy	Testing on different devices would help.
Risk Status	Identified

Risk ID	3
Risk Description	Fetching data sites structure change
Category	Requirements
Source	This was identified during early development and testing.
Probability	Low
Impact	Very High
Response	Accept
Strategy	Crawling the sites periodically.
Risk Status	Identified

5.3 PROJECT SCHEDULE

5.3.1 Project task set

Major Tasks in the Project stages are:

- Task 1: Ideation Phase.

- Task 2: Integrating various technologies and dividing the work.

- Task 3: APIs and Tools testing and usage.

- Task 4: Development of own modified or new tools.

- Task 5: Mobile Application Development.

5.3.2 Task network

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

5.3.3 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

5.4 TEAM ORGANIZATION

The project is being worked upon by a team of 7 people (2 project internal guide, 2 external guide and 3 project developers). Each project developer is aware how the entire project works. This is majorly possible due to the group size being small. Distribution of work has been done, now the tools available are being tested. It is decided to keep the team structure flexible throughout the project. Every project developer is subjected to equal contribution throughout the project, right from documentation to development. Refer table in Management reporting for further contribution details.

5.4.1 Management reporting and communication

Communication took place through mails and personal meetings and the tasks were monitored personally. The management was reported every once in 7 days.

Role	Participant	Responsiblity
Internal Project Guide	Yogesh Desai P.S. Vidap	1. Guidance for the project 2. Monitor the project plan. 3. Feedback and corrective action plans 4. Focus on team on project objectives.
External Guide	Akash Agrawal Vrushali Ingle	1. Guidance for the project 2. Feedback and Corrective Plans 3. Focus the team on project objectives.
Project Members	Ankit Bhagat Ayush Gandhi Manav Bharambe	1. Decide problem definition, requirement and risk a 2. Communicate process goals, status and progr 3. Assure quality of product that will meet the proje

CHAPTER 6

SOFTWARE REQUIREMENT SPECIFICATION (SRS IS TO BE PREPARED USING RELEVANT MATHEMATICS DERIVED AND SOFTWARE ENGG. INDICATORS IN ANNEX A AND B)

6.1 INTRODUCTION

6.1.1 Purpose and Scope of Document

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

6.1.2 Overview of responsibilities of Developer

What all activities carried out by developer?

6.2 USAGE SCENARIO

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

6.2.1 User profiles

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

6.2.2 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 6.1: Use Cases

6.2.3 Use Case View

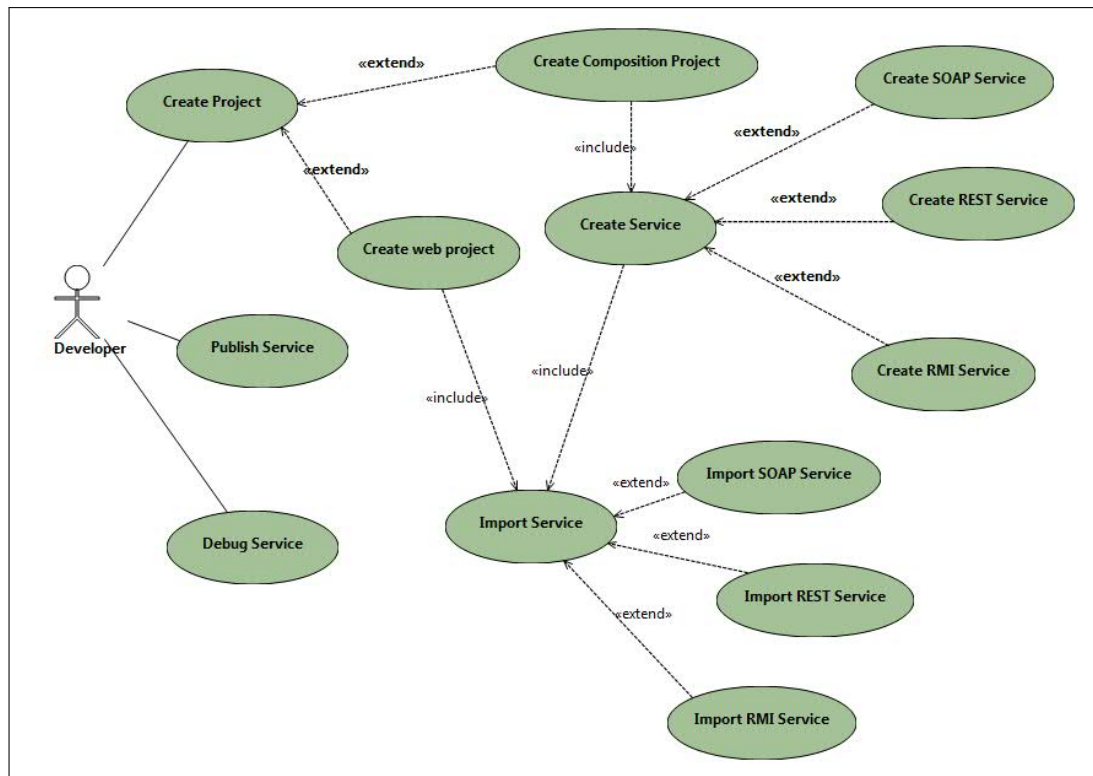Use Case Diagram. Example is given below

Figure 6.1: Use case diagram

6.3 DATA MODEL AND DESCRIPTION

6.3.1 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

6.3.2 Data objects and Relationships

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

6.4 FUNCTIONAL MODEL 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.

6.4.1 Data Flow Diagram

6.4.1.1 Level 0 Data Flow Diagram

6.4.1.2 Level 1 Data Flow Diagram

6.4.2 Description of functions

A description of each software function is presented. A processing narrative for function n is presented.(Steps)/ Activity Diagrams. For Example Refer 6.2
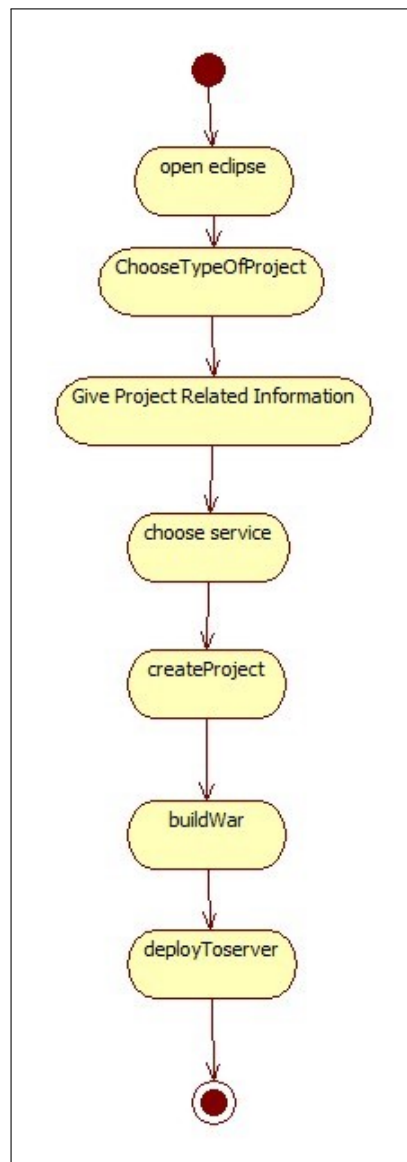

Figure 6.2: Activity diagram

6.4.3 Activity Diagram:

- The Activity diagram represents the steps taken.

6.4.4 Non Functional Requirements:

- Interface Requirements

- Performance Requirements

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

6.4.5 State Diagram:

State Transition Diagram

Fig.6.3 example shows the state transition diagram of Cloud SDK. The states are
represented in ovals and state of system gets changed when certain events occur.
The transitions from one state to the other are represented by arrows. The Figure
shows important states and events that occur while creating new project.

6.4.6 Design Constraints

Any design constraints that will impact the subsystem are noted.

6.4.7 Software Interface Description

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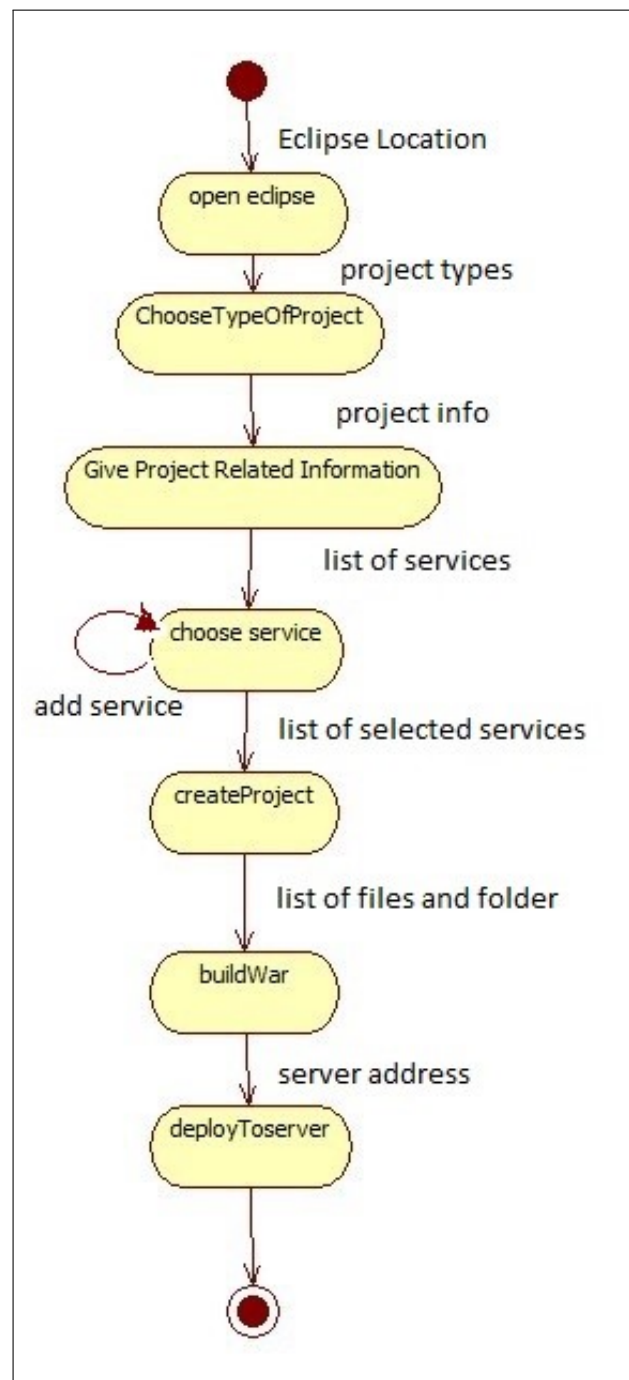interfaces to other devices/systems/networks/human are stated.

Figure 6.3: State transition diagram

CHAPTER 7

DETAILED DESIGN DOCUMENT USING APPENDIX A AND B

7.1 INTRODUCTION

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

7.2 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.
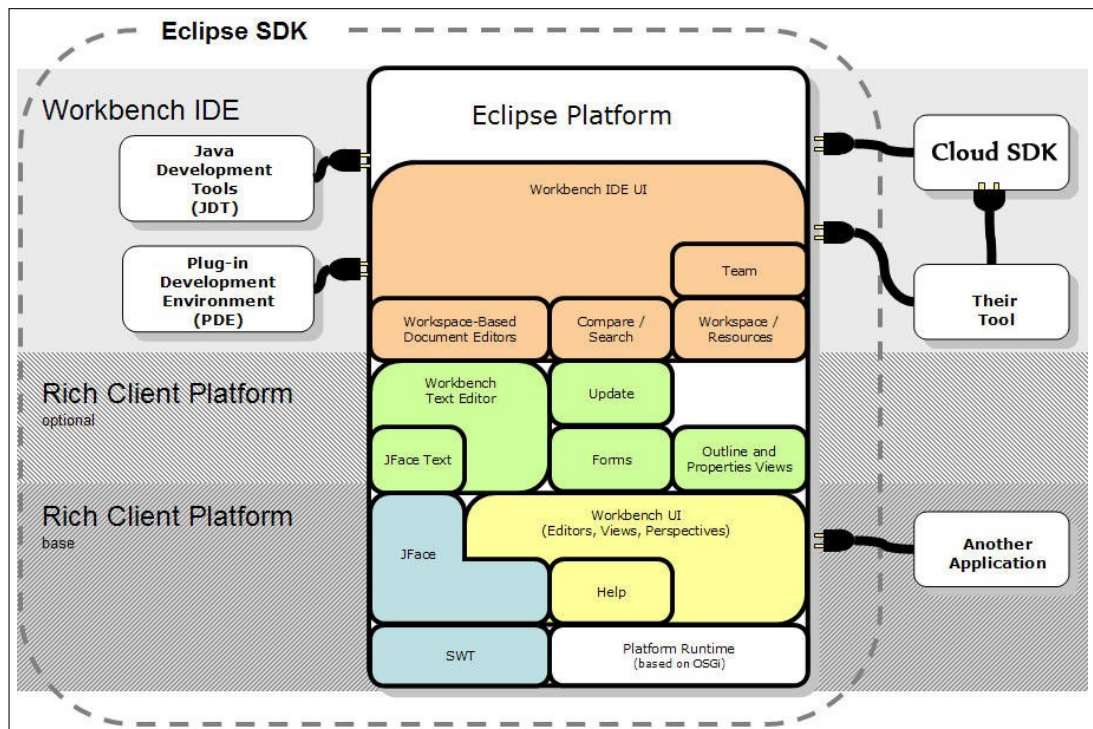

Figure 7.1: Architecture diagram

7.3 DATA DESIGN (USING APPENDICES A AND B)

A description of all data structures including internal, global, and temporary data structures, database design (tables), file formats.

7.3.1 Internal software data structure

Data structures that are passed among components the software are described.

7.3.2 Global data structure

Data structured that are available to major portions of the architecture are described.

7.3.3 Temporary data structure

Files created for interim use are described.

7.3.4 Database description

Database(s) / Files created/used as part of the application is(are) described.

7.4 COMPOENT DESIGN

Class diagrams, Interaction Diagrams, Algorithms. Description of each component description required.

7.4.1 Class Diagram

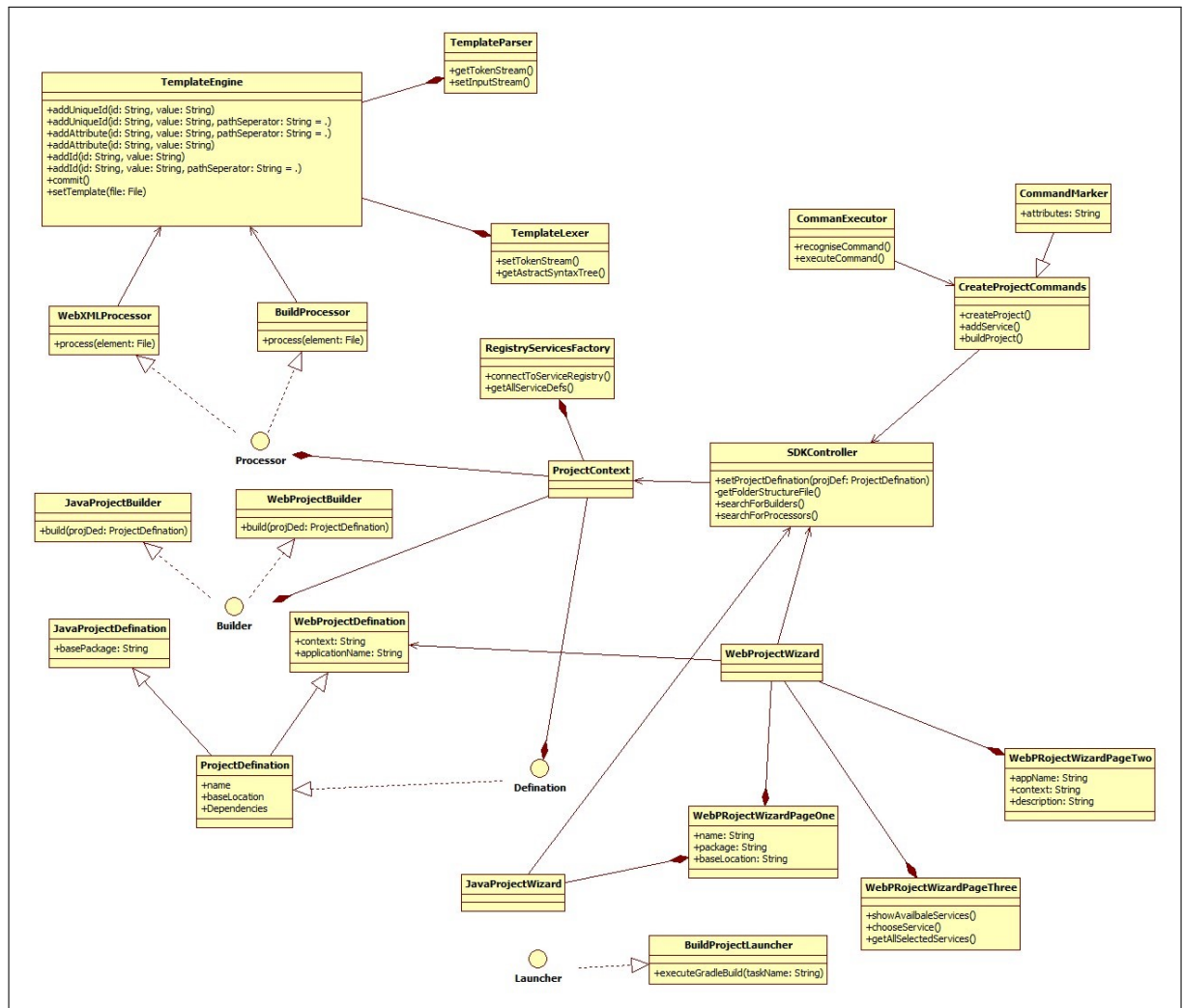Figure 7.2: Class Diagram

CHAPTER 8

SUMMARY AND CONCLUSION

Write one page summary and conclusion

ANNEXURE A

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 A.1: 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)

ANNEXURE B

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 [?],[?],[?], [?]

ANNEXURE C

PROJECT PLANNER

Using planner or alike project management tool.

ANNEXURE D

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 :

ANNEXURE E

PLAGIARISM REPORT

Plagiarism report