System Design

of

Best Buy Mobile Phone Using Opinion Analytics

Version 1.0

Prepared by

Zain Ali (FA14-BCS-045)

Semab Ali (FA14-BCS-018)

Supervised by

Mr. Tanveer Ahmed

Co-supervised by

Mr. Asif Muhammad

Department of Computer Science,

CIIT, Islamabad.

Dated: 14/12/2017

**Table of Contents**

[1 Introduction 1](#_Toc501012453)

[1.1 Purpose 1](#_Toc501012454)

[1.2 Scope 1](#_Toc501012455)

[1.2.1 Data Acquisition 2](#_Toc501012456)

[1.2.2 Pre-processing of Reviews 2](#_Toc501012457)

[1.2.3 Sentiment Analysis 2](#_Toc501012458)

[1.2.4 Visualization 3](#_Toc501012459)

[1.2.5 Product recommendation 3](#_Toc501012460)

[1.2.6 Seller information 3](#_Toc501012461)

[2 Design Methodology and software process model 3](#_Toc501012462)

[3 Data Representation 4](#_Toc501012463)

[4 Process Flow/Representation 6](#_Toc501012464)

[4.1 User Activity Diagram 6](#_Toc501012465)

[4.2 "Seller Crawler Activity Diagram 7](#_Toc501012466)

[4.3 Specs Crawler Activity Diagram 8](#_Toc501012467)

[4.4 Review Crawler Activity Diagram 9](#_Toc501012468)

[5 Design Models 10](#_Toc501012469)

[5.1 Context Level DFD: 10](#_Toc501012470)

[5.2 DFD Level 0: 10](#_Toc501012471)

[5.3 DFD level 1: 11](#_Toc501012472)

[6 Algorithm & Implementation 13](#_Toc501012473)

# Introduction

This software means to provide best product purchasing experience to its user. The main motive is to aware users from the product that they are purchasing. So they can buy a product comfortably and users get the best product which would be going to meet their requirements.

The selection of good mobile has been a serious concern of many people. Most of the people have not sufficient knowledge regarding purchase of a mobile. Users often have bad experiences of the product.

Our mobile recommender system will solve above problem efficiently and effectively. This system will allow its user to meet their requirements effectively. This system would provide searching option to meet user requirements in a better way.

This system would give the result of reviews based on sentimental analysis and let the users to know about the reviews of the people those are already using it.

This system will allow buyers to know the near position of seller. Buyers can contact to the sellers and purchase their desire product. Following are the modules:

## Purpose

The selection of good mobile has been a serious concern of many people. Most of the people have not sufficient knowledge regarding purchase of a mobile. Users often have bad experiences of the product.

## Scope

Our mobile recommender system will solve above problem efficiently and effectively. This system will allow its user to meet their requirements effectively. This system would provide searching option to meet user requirements in a better way.

This system would give the result of reviews based on sentimental analysis and let the users to know about the reviews of the people those are already using it.

This system will allow buyers to know the near position of seller. Buyers can contact to the sellers and purchase their desire product. Following are the modules:

### Data Acquisition

#### Web Spider/Crawler

We will create our own web crawlers that would target to e-commerce website to gather reviews

#### Text Cleansing

In text cleansing we would remove the HTML Tags from the text to get to our desire result.

#### Information Extraction

#### Here, we will target to review section of the web pages.

#### Database storage

The reviews would be stored into a database.

### Pre-processing of Reviews

#### Sentence splitter

Each review consists of multiple sentences. Here we will split each review into single sentences.

#### Tokenization

The single sentence would now converted into tokens.

#### Normalization

Here we will normalized the reviews

#### Punctuation Removal

All the punctuation in the review would be removed in this step.

#### POS tagging

In this step each word would be marked as its proposition.

#### Stopping word Removal

Here all the useless words would be removed.

#### Stemming

Each word would be converted to its root word

#### Lemmatization

### Sentiment Analysis

#### Machine learning

Naive Bayes algorithm would be used for review sentiment analysis.

#### Stanford Dictionary

Dictionary that contains word scores.

#### Dual Prediction

After sentiment analysis review would be classified as negative or positive.

#### Feature Rating

The features in reviews of the product being considered would be rated at this point.

### Visualization

#### D3.JS

This API would be used to visualize the information to users.

#### Three.js Web graphics would be manage by this API.

### Product recommendation

#### Feature Filters

Users would be able to filter out the product on based on their own requirements.

#### Aspect Based Features recommender

### Seller information

#### Contact Information

Buyer will get the contact information of the seller.

#### Seller location

Buyer can meet to seller at the location where he would be.

# Design Methodology and software process model

The design methodology that we used in our system is Procedural. Our system is basically a recommendation system and we have incorporated machine learning in it so that it recommends more efficiently and in a realistic manner.

**Justification:**

The system have to interact with other websites to scrap data from them i.e. reviews and mobiles model from official websites of Company’s.

Therefore we use python to do above all these things because python is simple, elegant and consistent and also its learning curve is small. We use procedural approach in it the reason is there are not many independent task in our system and there is not too much reusability we need because every website have different structure so we have to develop a unique crawler to handle it and also our system has complex backend computation so we are using procedural approach to tackle this situation, procedural is more efficient and increased performance than OOP approach.

We use waterfall process model for the development because Tasks flow down the list in sequential order and we have to show to project committee each phase to know whether project is on right path or not. And the reason behind its usage is that requirements of our system is clear and it is not expected to change and Tools and technology is also known and well understood. The waterfall also allows us early design changes when the specification documents in the first couple stages fleshes to the project committee and the scope is not changed too rapidly.

# Data Representation

There are 5 tables in our database. Each table has attributes shown in following diagram. Brand table contains name of mobile brands e.g. Qmobile, Samsung, Oppo etc. Brand table has a relation with “Mobiles” table that is one to many. Review table contains reviews of each models of mobile. The score table contains the “sentiment analysis” and “Naive Bayes” result of each mobile stored in mobiles table. The comment table is used to store comments of each model.

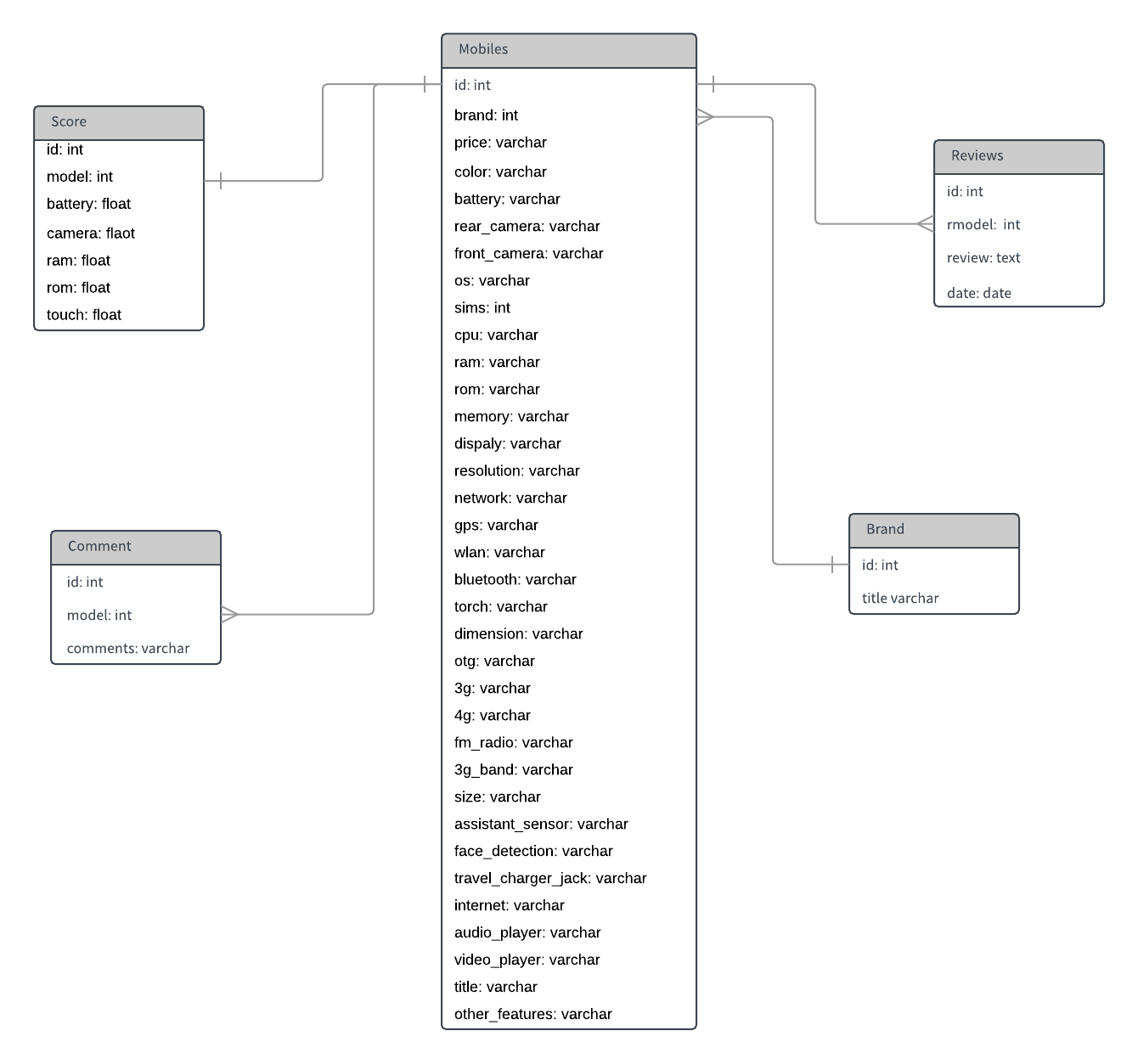


Figure Entity Relation Diagram

# Process Flow/Representation

## User Activity Diagram

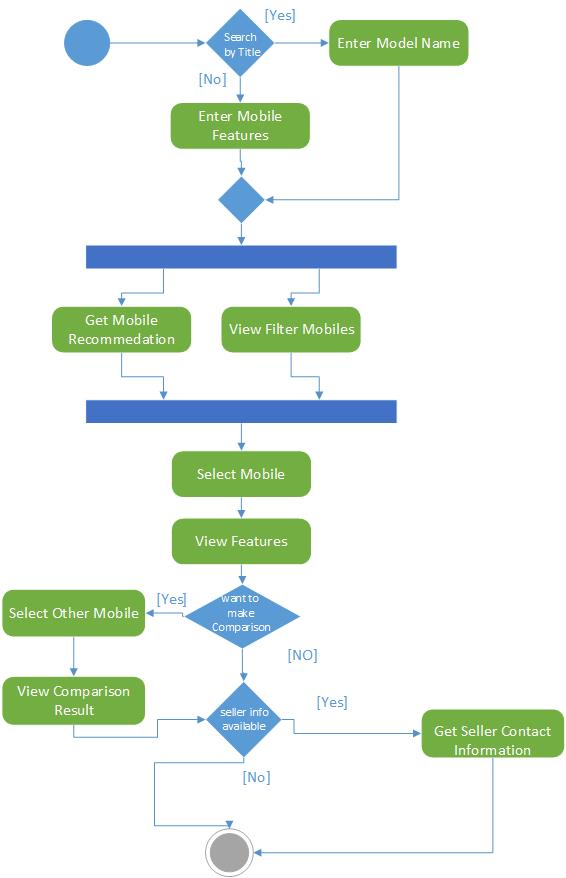


Figure “User” Activity Diagram

## Seller Crawler Activity Diagram

Figure “Seller Crawler” Activity Diagram

## Specs Crawler Activity Diagram

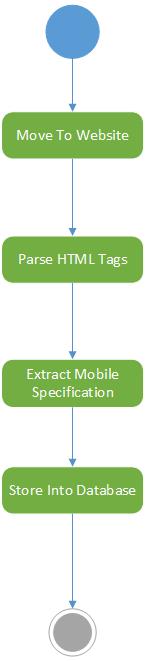


Figure “Specs Crawler” Activity Diagram

## Review Crawler Activity Diagram



Figure “Review Crawler” Activity Diagram

# Design Models

## Context Level DFD:

User can perform multiple actions i.e. view mobiles and can request for recommendations from the system. System can request crawler for multiple services like bringing the reviews and seller information.

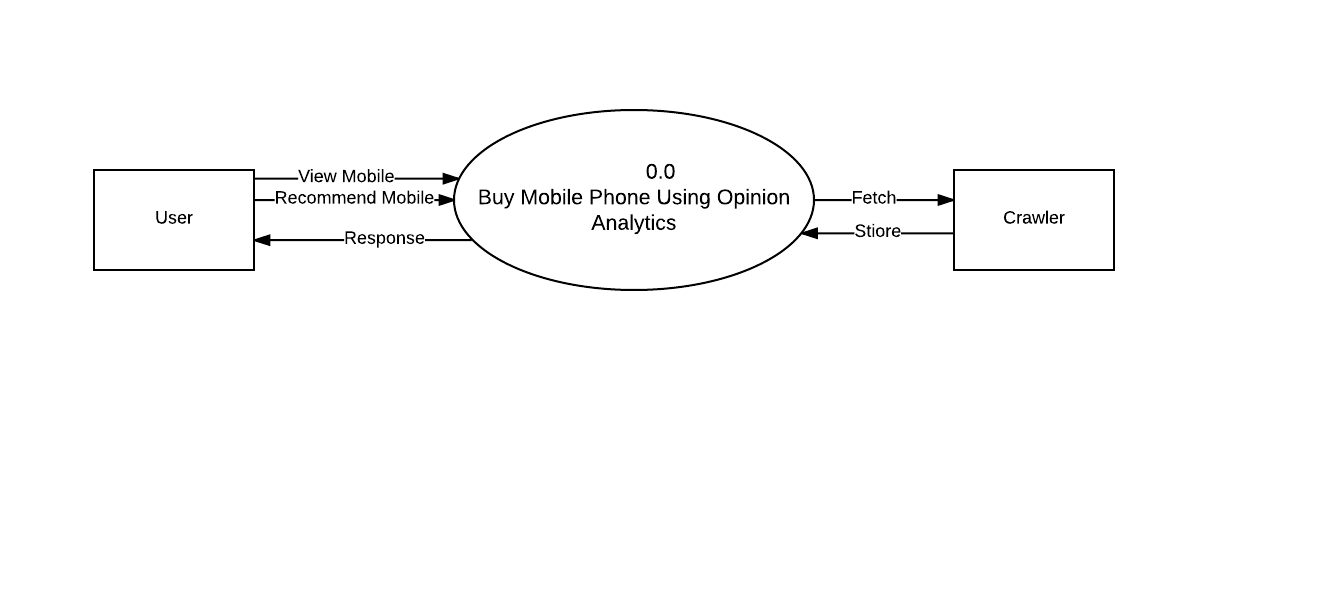
****

Figure : Context Level DFD

## DFD Level 0:

In sentiment analysis, sentiment scoring will be calculated which passes the result to recommendation which calculates overall score and which passes the result to visualization for showing the result to user.

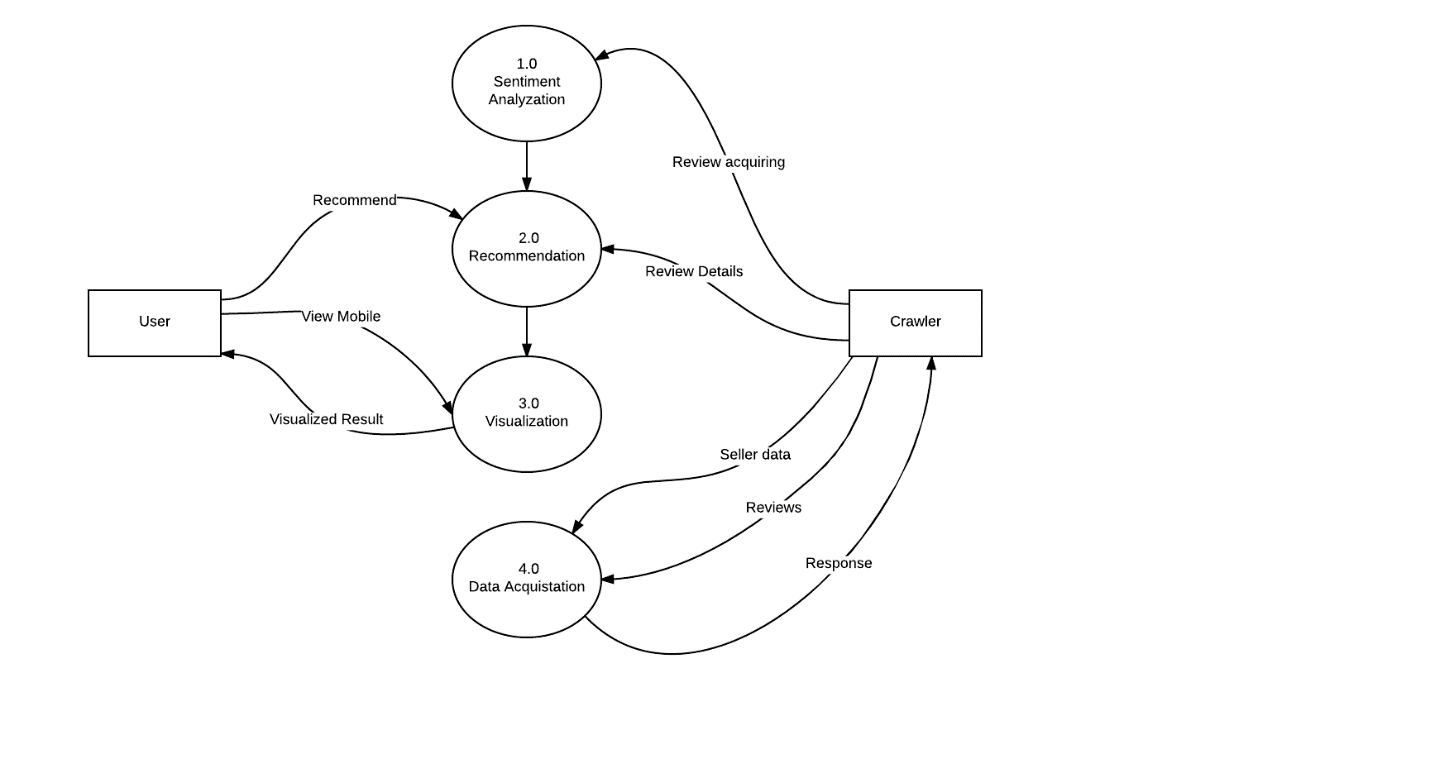


Figure : Level 0 DFD

## DFD level 1:

In this level Sentiment analysis is further break down to tokenization, stopword removal, lemmatization and sentiment scoring to calculate the sentiment analysis and recommendation is further broke down to validate request and get recommendation which calculates final score. Seller info is acquired by visiting the website then parsing the data and fetching the required information from parsed webpage.

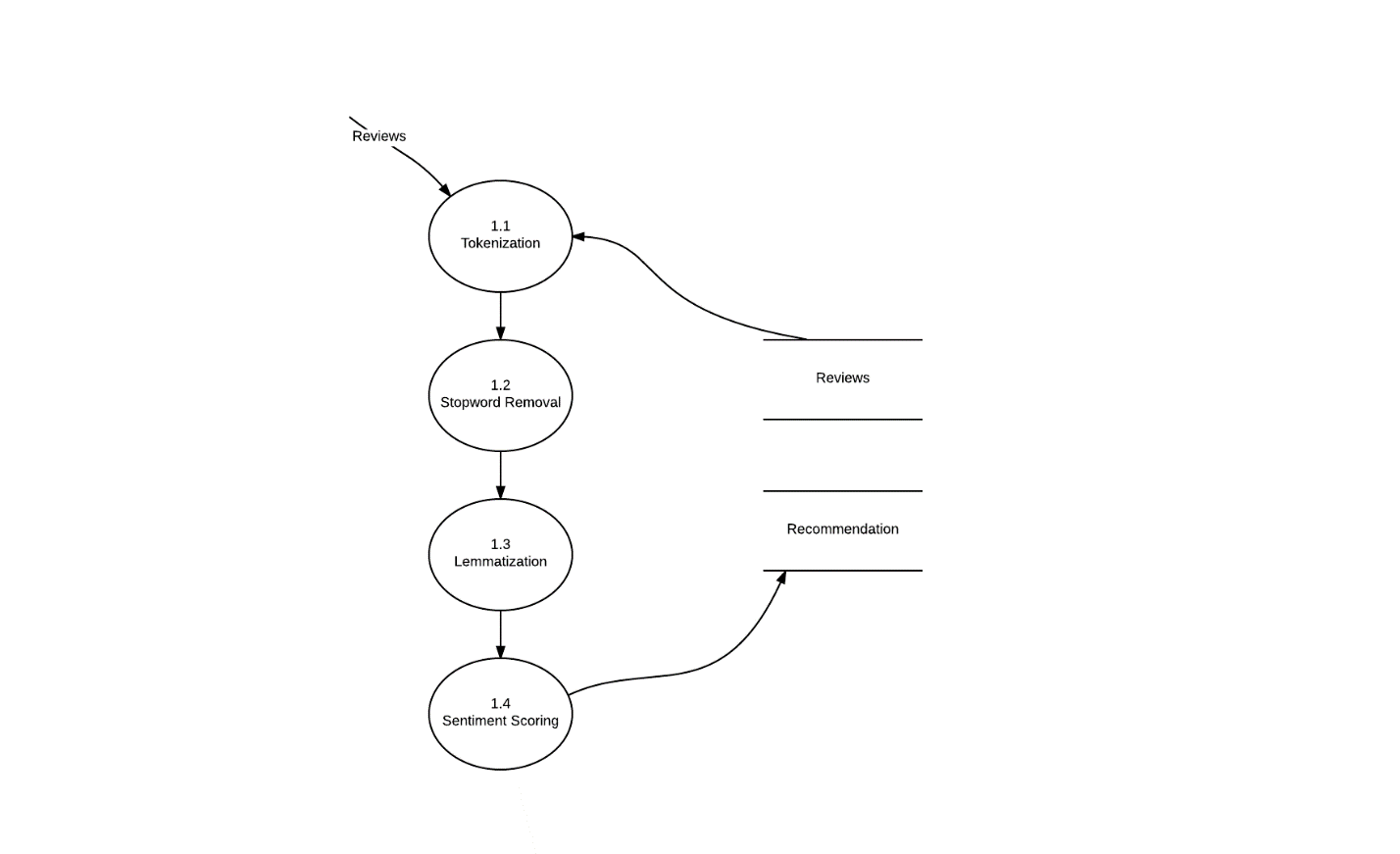


Figure : Level 1 DFD

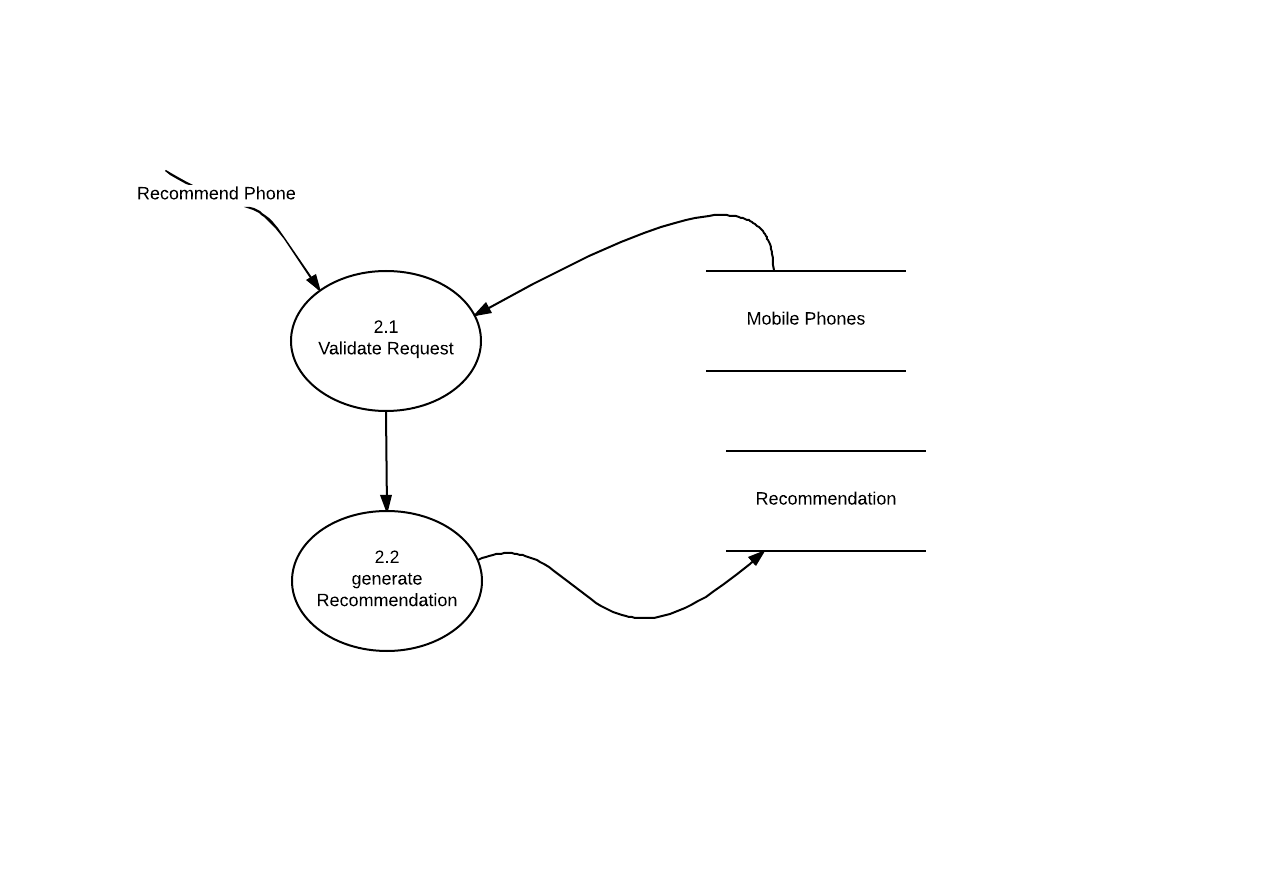


Figure : Level 1 DFD

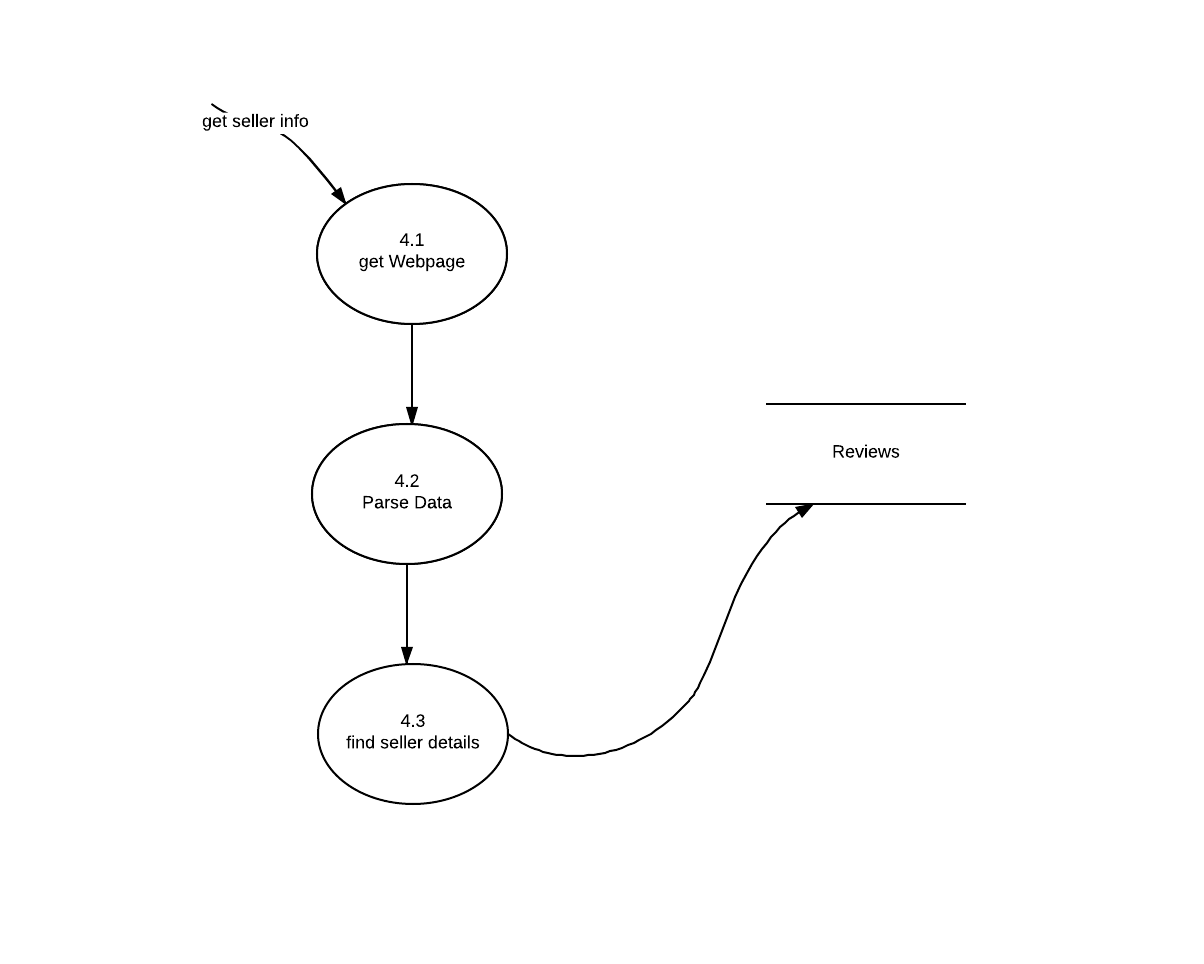


Figure : Level 1 DFD

# Algorithm & Implementation

**Data Acquisition:**

1. Initialize the variable i.e. the name of website to be scrapped and with query
2. Get all information complete webpage using request.get() method
3. The response is then decode in using parser in html format
4. The div containing links to respective pages is retrieved
5. Append the sub links with main links
6. Now visit each link with the help of loop
7. Retrieve the user reviews by targeting the div that contain the reviews
8. Remove html tags and unnecessary information from retrieve data and store in list
9. Then make connection with database
10. Traverse each review in list and store them in database.

**Pre-processing (Dataset Preparation):**

1. Load the reviews in from database in List
2. For each review retrieve from database the following steps are performed
3. Tokenize the review in sentences then into words
4. Remove the punctuation and stop words from each review
5. Perform lemmitazion on words for better sentiment score
6. Calculate sentiment score for each review
7. If sentiment score is greater than 0 than review label as 1(means pos) and 0 (Neg)
8. Make a csv file and store it a review and sentiment score