Bilkent University

Department of Computer Engineering

Senior Design Project

Brandlyser: Web Based Analyzes Systems

High-Level Design Report

Ali Şengül, Ferdi Aslan, Sera Büyükgöz, Tunahan Kunt

Supervisor: Hakan Ferhatosmanoğlu

Jury Members: Erman Ayday and Özgür Ulusoy

Web Page: www.tunahankuntt.wix.com/brandlyser

High-Level Design Report
Dec 24, 2015

This report is submitted to the Department of Computer Engineering of Bilkent University inpartial fulfilment of the requirements of the Senior Design Project course CS491

Contents

List of Figures

Analysis Report

Brandlyser: Web Based analyzes Systems

1 Introduction

In today's world, gathering information on internet before doing an activity becomes popular and reliable. Almost every brand in the industry need to be informed about their reputation. However, every day people complain different parts of their life to the world through internet such as social media, blogs, websites etc. They complain about where they go, the quality of the product that they buy, their feelings and opinion about almost every topic. They also comment each other's complaints. That creates a huge amount of data on the internet. That huge amount of information awaits and ignores because of mess on the internet. Since that data is not classified or listed. To reach information about reputation of brand, that data need to be collected and classified according to the specific topic to be use in the industry.

Automobile industry is one of the topic that people mostly talk about on the internet. Brandlyser aims for gathering information about that brand and that brand's model via using Twitter and Facebook posts, using comments under the commercial videos and related other videos in YouTube, posts on the some popular Blogs and some complaints Blogs. Program synthesis that data then gives analyzes with statistic of market place of that brand on the internet.

1.1 Purpose of the System

Brandlyser aim is to create an analysis system. That system provides brands' companies statistical analysis of customer's opinion about product by using comments on websites, blogs, social media etc. Companies can enter the system to see their analysis results. They also ask future analysis or change default settings of analysis by adding more websites or features. Request page will lead company about future plans and inform them by sending mail. In that way, company does not need to contact with providers. The other side of the medallion, provider enter brand's information, keywords and chosen websites and features by easily using brand's definition page. The cost of analysis determined according to the size of the data using throughout the analysis. In the first step, size of the data are determined then cost calculated by using a formula which includes number of websites, detailed features, size of database which is using in comparison algorithm. Then, provider prepares an username and password for company. System starts to work with gathering comments to the websites and create a data set for comparison algorithm. Comparison algorithm work and turn statistical results to the brand's dash box. When results are ready, company inform by mail. Finally, company can reach their results by using their username and password.

1.2 Design Goals

Reliability of Content: The purpose of our system is give reliable analysis to the customers. We need to examine as much comment as possible and provide good and reliable analysis according to these comments. This means that our algorithm should choose usable comments and remove the noise in big data.

Response time: Our system will review thousands of comments and provide a report according to these comments. Processing this big data will be time costly so we need to improve our system so that this algorithm works in minimum amount of time. Also we will create a website for users to find their results. This functions need to take small amount of time for better usage of the system.

Modifiability: The purpose of our system is providing reliable results. So we need to improve our algorithm as much as possible and our system should work properly after this changes.

User-friendliness: The web-site that we allow users to start a search and see the result of the search should be user friendly. The graphical interface should be easy to understand and use.

Accessibility: The website should be accessible from computers and mobile devices that support the graphical interfaces. Customers should be able to access the website anytime they want. If the system will be offline due to upgrade, customers will be informed.

Security: Since our customers will pay for our service using our web-site, the web-site should be safe. System shouldn't keep any credit card in information. Interaction between system and bank shouldn't be accessible by anyone.

1.3 Definitions, Acronyms and Abbreviations

Brandlyser: Name of the system which is created.

Providers: People who controls and operate service of system, basically us.

Features: Different properties which company chooses to specify analysis.

Websites: Systems source of data such as; YouTube, Twitter, Instagram, Blogs etc.

Dash box: Results page which consist of little boxes that leads specifically detailed analysis.

1.4 Overview

Brandlyser will aim to be a branch of the company on the web. Brandlyser collects data which is related to the brand through the web and store them separately in Brandlyser's own database. Brandlyser will only work for automobile industry. The goal is to use Brandlyser in almost every industry whose products have value on the internet. Almost all brands use Brandlyser to follow popularity of their product through the web. Analysis will be more detailed when company decide to use many features. In order to reach these goals, Brandlyser need to be more complex algorithms. The system need to be support all of these data. First, we need to create a database to cope with storing all the data from collecting through the Internet. Second step is to create space for system to analysing data. So we need to cope with parallel computing algorithms to link some processors for creating space to the system. Final step is to update user interfaces. Since our customer's request is advanced, we need outstanding service interface. To addition to all of them, interactions between layers and subsystem must be efficient.

2 Current Software Architecture

Figure 1: BoomSonar User Interface [1]

Despite the fact that there is no software architecture gave in any web and social media monitoring tools, certain resemblances and contrasts between Brandlyzer and current systems can be characterized. As an example, Radian6 and Boomsonar both listen, share, learn, and engage with entire social web. They are communicating via a news site, discussion forum, blog, or a micro blog like Twitter. They absorb all this information and interpreting them to provide a database system that stores information for their customers. Radian6 enables its customers to track, monitor and react to comments and questions according to demography, trends, and targets. Boomsonar provides social media following, measurement, and reporting and online prestige management to its customers. It can provide localized results for every country. So Boomsonar and Radian6 have an interpretation sub-system in their architecture.

3 Proposed Software Architecture

3.1 Overview

Brandlyzer will be implemented by different layer. We will have actually 3 main parts. So we will use 3-tier system decomposition. First one is about collecting data from different web sites such as youtube, instagram and blogs. Second one is responsible to analyze data and the last one's responsibility is to present data result. By 3-tier system, our software will be more scalable and secure.

Data collector: This layer is a subsystem which will collect categorized data from web. Categorized means that this layer gets comment from different web sites and if the commented brands are same, data collector will keep them in related area in its memory.

Data analyzing: This layer's main activity is to run the analyzing algorithm on comments. It gets every comment from data collectors and after analyzing them, it gives the categorized result to presentation layer.

Presentation layer: This layer's main activity is to contact brandlyzer's users. It gets result from data analyzing layer and show results to customers. It will be a single web site which has only a few attributes.

3.2 Subsystem Decomposition

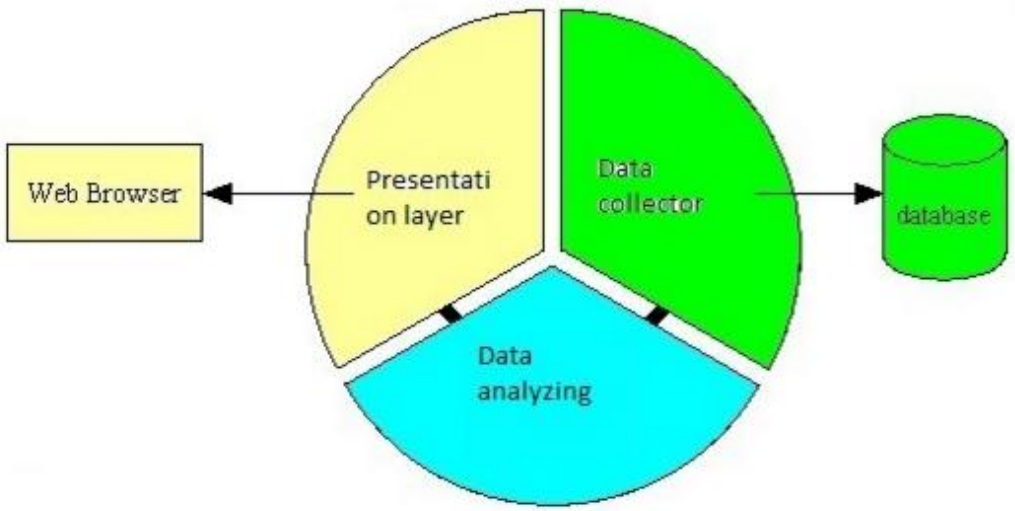

Figure 3: Subsystem Decomposition

As we see in the figure 1, we have mainly 3 layer. Data collector, data analyzing and presentation layer. More detail will be explained below.

3.2.1 Data Collector

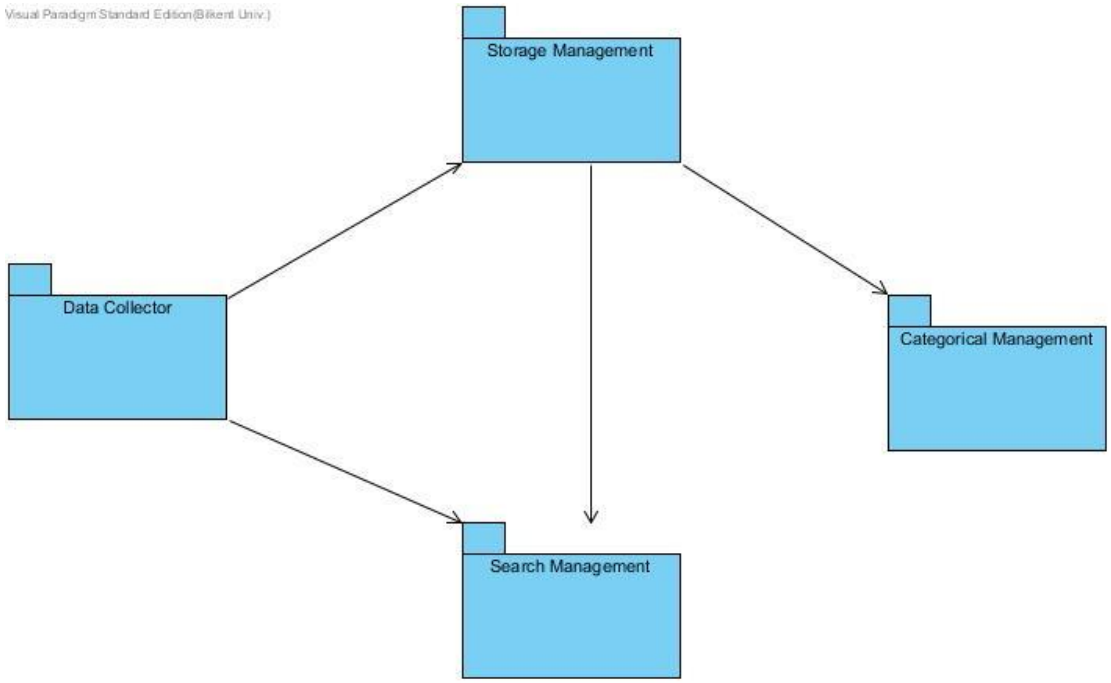

Figure 4: Subsystem Decomposition of Data Collector Layer

Data collector layer's main activity is to collect data from web sites.

- ***Storage management***: This subsystem keep collected data by categorized style.
- ***Search management:*** Searching every related data from web sites is belong to search management. It search every comment under analyzing brand.
- ***Categorical management:*** It categorize comment in terms of related brand. For example, from 2 web sites search management can find comment about the same brand. In this issue categorical management issue is keeping these comment related blocks in storage management.

3.2.2 Data Analyzing

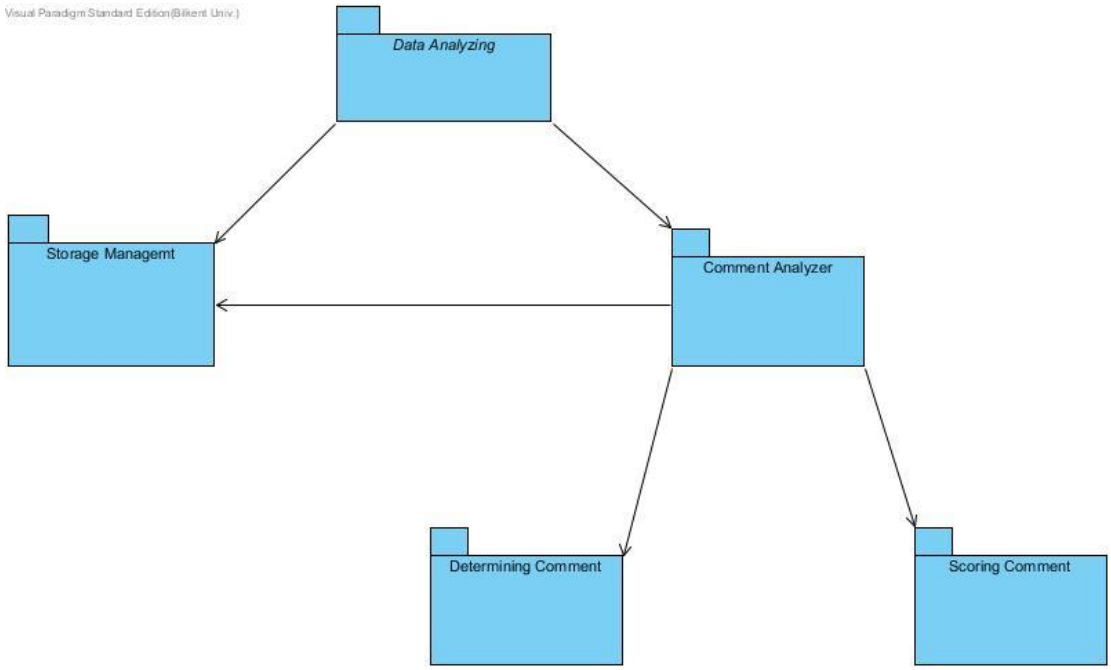

Figure 5: Subsystem Decomposition of Data Analyzing Layer

This layer analyzes comment in data collector.

- ***Storage management :*** It keeps analyzed comment
- ***Comment analyzer:*** It has to analyze collected comment from web sites. It has 2 main activities. Determining comment, Scoring comment
- ***Determining comment:*** We will have lots of comment and all of them are not analyzable. In order to get analyzable ones, determining comment subsystem will very helpful for our software.
- ***Scoring comment:*** Every analyzable comment will be analyzed in this subsystem. Our analyzing algorithm will work in this area. After analyzing, algorithm will give score to comment in terms of positive score or negative score. These results will be kept in storage management.

3.2.3 Presentation Layer

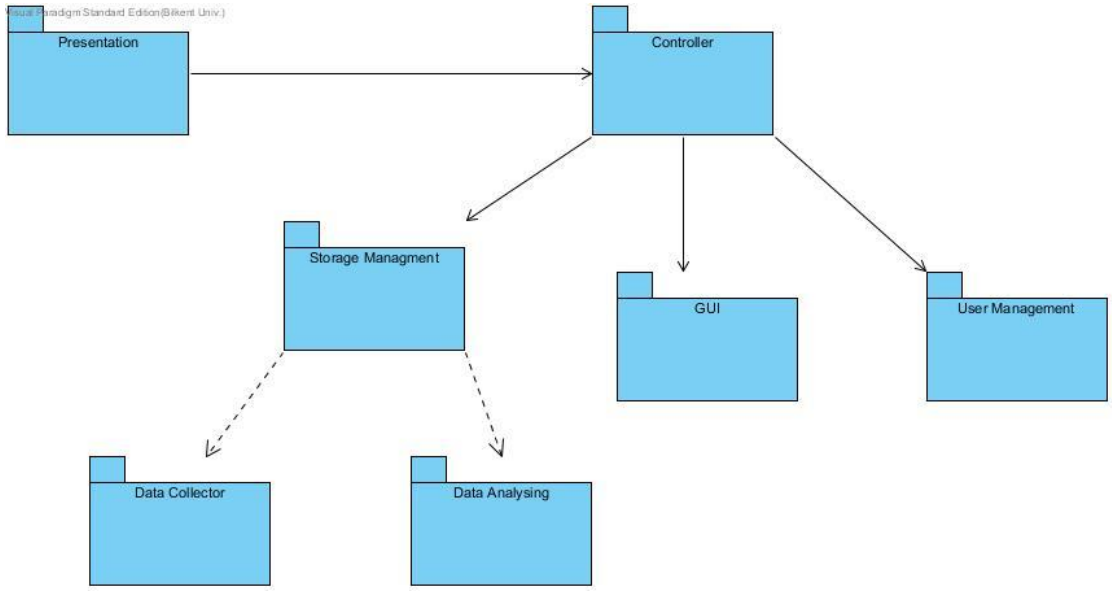

Figure 6: Subsystem Decomposition of Presentation Layer

- ***Controller:*** General controller for presentation layer.
- ***Storage management:*** It gets result from data analyzing and it can also get size of data from data collector in order to show size of analyzed data to user.
- ***GUI:*** It will have a easy system. First screen has a little boxes and every box belongs to a web site. In this box, every website has their general analyze. Clicking these boxes opens new page and detailed results are shown.
- ***User management:*** It keeps user information and enable user to login and logout.

3.3 Hardware – Software Mapping

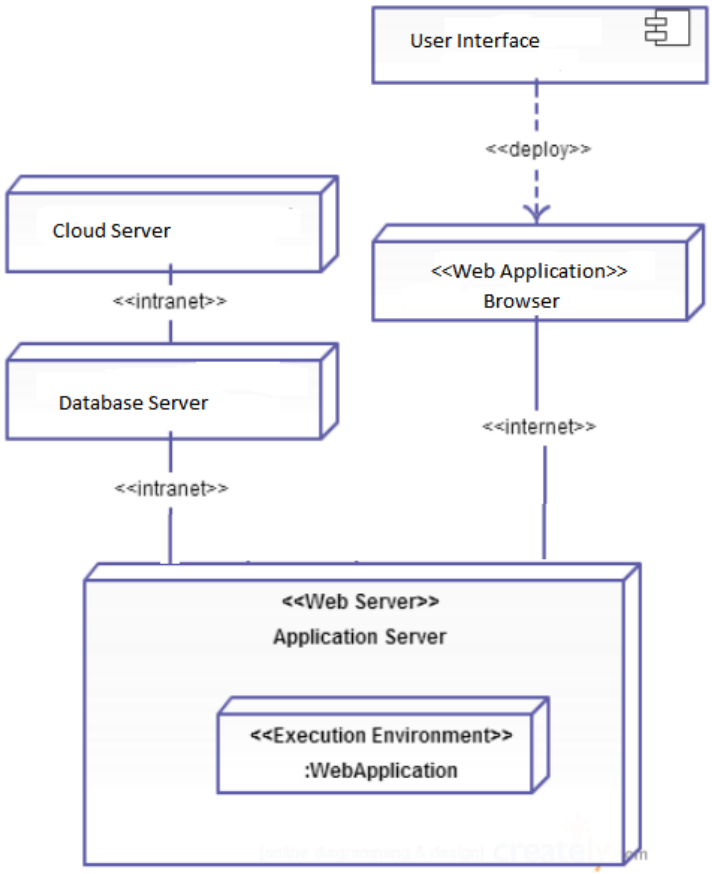

Figure 7: Hardware – Software Mapping of Brandlyser

Hardware/Software mapping aims to map between software components and connectors of the system and the hardware part of the application on which the software executes. Brandlyser requires a server that dash boxes always visible to customer. Moreover, algorithm works on that server as and background programme User can reach their dash box result by using a device has an internet browser and connection. These devices should be connected to Brandlyser's cloud system via internet. All data of Brandlyser will be held in cloud server.Application server has one to one intranet connection with database service which is cloud server.

3.4 Persistent Data Management

We are planning to use mongoDB to keep sentences in database, as mongoDB is efficient for single data.

Our database generally store sentences. Firstly, determining comment subsystem will do its occupation and find the analyzable sentences. There will be another row for analyzable or not? After that, scoring comment subsystem will handle its job and score sentences in terms of positive and negative meanings. There will be another row for this attitude.

For user information there will be another query to keep. This information is not a big data. So to keep them is very easy but the problem is security. That is because there will be a competition between brands and if there is a security problem, it will be harmful for software reliability.

3.5 Access Control and Security

In Brandlyser, every client that is signed in to the application can use all functions of the application. While registering, software clients will take their username and password. They will also enter their email address for activate their profiles. By clicking relevant link which will be send to their e-mail address as a confirmation email, clients will have an active profile. When they want to use software, they have to enter correct username and password combination. When they lost their pass they need to ask for a new password us.

For security, every client's data will be stored in database. Client's' password will be encoded with cryptographic algorithms to avoid passwords being stolen. Also to prevent SQL injection to database, each query will be filetered. Lastly, data communication between software and client will be carried out by secure HTTPS.

3.6 Global Software Control

Data communication between software and client will be carried out HTTPS. To prevent losing of any data, concurrency issue will be handled. For concurrency MongoDB have a solution. MongoDB “uses multi-granularity locking that allows operations to lock at the global, database or collection level, and allows for individual storage engines to implement their own concurrency control below the collection”[3].

3.7 Boundary Condition

Initialization: Clients need an internet connection to startup. Then they need to open their web browsers and navigate to our web site. To start to see their reports they need to log in to their accounts.

Termination: If clients close their web browser or computer or they log out from their accounts application will be automatically terminated.

Failure: If any unexpected something happened in the server the system will give an error and no one can use the application. On the other hand, to decrease the probability of errors in the server we will be purchase reliable server. At the side of clients, if they do not have internet connection, the system will throw error to clients.

4 Subsystem Services

Presentation layer: Presentation layer is responsible for visualizing. The interaction between system and users will be handled by this layer. All the functions (start search, view result, login, logout...etc) can be accessed in this layer. User management keeps the information about users and it will engaged to system whenever user-related function are needed. GUI is responsible for visualizing the user interface. It will put the necessary information to related areas. Store management keeps the information about data. This data contains the amount of comments algorithm analyze and what is result of this analyzes.

Data Analyzing: This layer is responsible for analyzing collected data. It gets store comment from data collector. Every comment must be determined by determining comment subsystem. This subsystem gives the analyzable comment to scoring comment. The main algorithm will be run in this subsystem and give result to storage management. After analyzing a comment, computer will get the idea in terms of positive and negative meaning. And finally, it gives results to storage management.

Data Collector: Data collector will collect data from web sites and by the subsystem categorical management; it will keep comment by categorized way. By this way, data analyzing layer will get the related comment and make the analyzing easy for computer. The categorization works depends on brand. Every comment is searched by search management and given to categorical management. After finding comment categorical management gets the brand name and find the related one in storage management.

5 Glossary

HTTPS : Secure hypertext transfer protocol

MongoDB: Cross-platform document-oriented database. Classified as a NoSQL database. [4]

6 References

[1] BoomSonar, "sosyal medya takip | raporlama | analiz | yönetim | ölçümleme", 2015. [Online]. Available: http://www.boomsonar.com/tr/. [Accessed: 24- Dec- 2015].

[2] Salesforce Marketing Cloud, "Products: Social | Radian6", 2015. [Online]. Available: http://www.exacttarget.com/products/social-media-marketing/radian6. [Accessed: 24- Dec- 2015].

[3] Docs.mongodb.org, "FAQ: Concurrency — MongoDB Manual 3.0", 2015. [Online]. Available: https://docs.mongodb.org/v3.0/faq/concurrency/. [Accessed: 24- Dec- 2015].

[4]SearchDataManagement, "What is MongoDB? - Definition from WhatIs.com", 2015. [Online]. Available: http://searchdatamanagement.techtarget.com/definition/MongoDB. [Accessed: 24- Dec- 2015].