Software Engineering I Group II

Health Monitoring Analytics 2014

D. Yao, W. Fang, W. Zhang, Y. Sun, Y. Wu, Z. Zheng

https://sites.google.com/site/healthmonitoringanalytics2014 Oct. 14th, 2014

Individual Contributions Breakdown

“All team members contributed equally”

Table of Contents

1 Customer Statement of Requirement

1.1 Problem Statement

1.1.1 Health monitoring analytics with Twitter

With the development of Social Networking Service (SNS), Health Care Analytics [1] (HCA) discovers a new world. Traditionally, HCA is based on Personal Activity Monitoring (PAM) which collects individual's activity data with sophisticated carry-on devices, e.g., Google Shoe, Nike Fuel, Zeo Sleep, Jawbone's Up, etc. In PAM, customers could adjust their excising intensity depending on the analyzing results by comparing their activity data with the standard health statistics that come from relevant health researches. Although some PAM devices with network connection could help customers share their activity data and analyzing results with others, there is still a limitation of user population. In another word, "none of these products is able to provide analytics for tracking population activities related to healthy lifestyle and gaining insight into lifestyle trends" (see ref. [2]). However, this is what the customers need – the customers are not satisfied with merely looking at their own activity data and analyzing results, and they are also eager to see that of others from all over the world. This is proved by more than one in three American adults who go online to figure out a medical condition surveyed by the Pew Research Center's Internet & American Life Project in 2013 [3].

Due to the emergence of the powerful SNSs with large amount of users and their information, for example, Facebook, Google+, Twitter, etc., it is possible for us to gain public data relating to health issues in order to meet the customers' requirement. Figure 1-1 shows the number of monthly active users in different SNSs in 2014 [4]. Although Facebook and Google+ seem to have more monthly active users than Twitter, each user's information cannot be accessed without his or her authentication because of the privacy, which means we are unable to crawl the private information of the public from Facebook and Google+ database. On the other hand, Twitter is much more friendly. We could crawl the past and real-time Twitter data stream provided that we register a Twitter developer account. Thus, we chose Twitter database as our data source. The goal of our team is to mine and analyze the online information about people's exercising activity and other health related issues, to show useful public analyzing results and also customer-oriented health suggestions [2] by using Twitter Application Programming Interface (API).

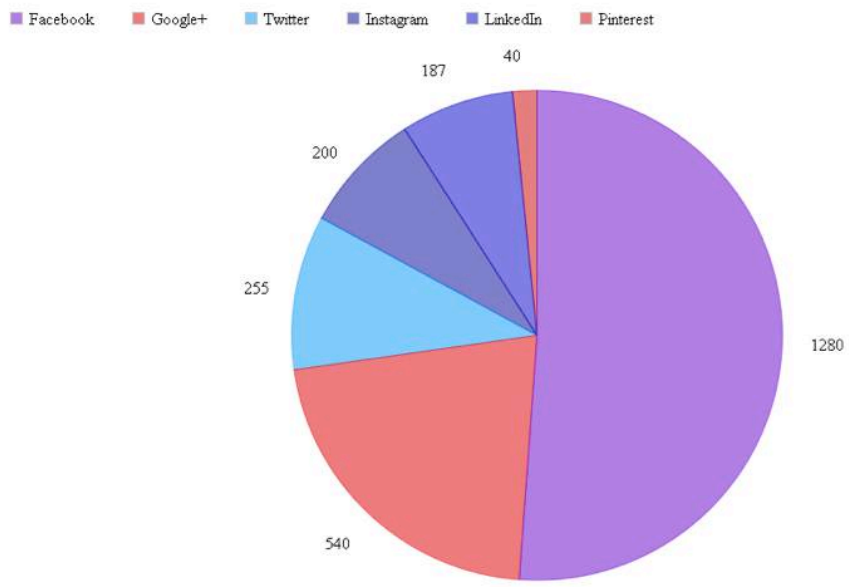


Figure 1-1. Monthly active users (million)

1.1.2 Existing works

Based on the Twitter API, three former groups have developed their own Health Care System (HCS). Although they use different third party APIs for Twitter development - group 1 takes Phirehouse [5] while group 2 and group 3 take Tweet Tracker [6], the scope of the data (always in a JSON format [7]) which they could scrawl from Twitter database is the same since the same Twitter access rules - Rest API and Streaming API [8]. The significant difference among these groups is how they use these data, i.e., what are their features. The comparison of features among the three groups is shown in the Table 1-1.

Features	Group 1	Group 2	Group 3
Tweet heat in geographical distribution (heat map)	√	√	√
Tweet heat in geographical distribution (marker map)	√		√
User ranking		√	√
User Profile			√
Find Partner		√	

Table 1-1. Comparison of features among the former groups

Tweet heat is calculated by counting the number of tweets. To improve the accuracy, group 2 employs two methods: combination keywords and keywords weight. Although they show the distribution of tweet heat in heat map and marker map, the display cannot be varied by time or exercising type. Besides, the heat map is crude, for example, in the group 3's android application, the heat map is simply displayed by implementing the circle overlays on the Google Map – no gradient. The user ranking is also based on tweets counting and the display cannot be varied by the state, time or exercising type. The user profile in group 3's android application shows the user's tweet history, however, there is no combination between the user's data and the public data. As for the finding partner, it is a special feature in group 2, but the finding operation should be in a much more intelligent way. In conclusion, all three groups obviously lack sufficient analysis from Twitter data, probably because they spent too much time on establishing the system and cannot figure out some technique problems. Thus, our team is going to pay more effort on analysis. The Table 1-2 shows the display platform in three groups.

Platform	Group 1	Group 2	Group 3
Website	√	√	√
Android		√	√

Table 1-2. Platforms of the former groups

1.1.3 Improvement and extension

Our team takes the advantage of the existing data collection infrastructure from group 1 because we prefer to use the Phirehouse. Phirehouse is the Twitter streaming API which can download the real-time data from twitter database. However, we also need past tweets which are not used by group 1, so we ourselves find a good Rest API on GitHub [9]. Our team also takes the advantage of the combination keywords and keywords weight method from group 2. But we have refined the keywords settings, for example, we make it equal among exercise type keywords, so it is more reasonable when comparing the exercise type. Besides, we add user demography information in the user table such as age, gender and type (people or company) by text analyzing [10]. Unfortunately, the accuracy of this demographic speculation is not shown in their website.

Features	Description
Tweet heat in geographical distribution	Counting the tweet number in different states shown as a state map which can be

	varied by time, exercising type and demography. Shown as bar charts. Shown as a heat map with gradient. Shown as a marker map with screen name and tweet text.
Tweet heat in variation tendency	Counting the tweet number in different day in a week, and different time in a day which can be varied by state, exercising type and demography. Shown as line charts.
Tweet heat in exercising classification	Counting the tweet number in different exercising type which can be varied by state, time and demography. Shown as pie charts.
Tweet heat in demography	Counting the tweet number in different age and gender which can be varied by state, time and exercising type. Shown as pie charts.
User ranking	Ranking the user by number of tweets which can be varied by state and exercising type. Shown as tables.
Exercising duration	Extract the exercising time in the tweet text, figure out the mean, max and std. dev. in different state and exercising type and demography. Shown as tables and a state map and a marker map which present the screen name, exercising time, exercising type around you.
Exercising frequency	Figure out the mean of how many days in a week that different users exercise in different state and exercising type and demography shown as bar charts, and as a state map.
Word frequency	Split the words in tweets and count to find out related topics for correlation topics analytics. Shown as a table.
Correlation topics	Draw the linear regression among different health topics in line charts to find out how close the relationship is.
Tweet sentiment	Match the tweets with different mood and count the number in different state, type, time and demography.
Personal diagnosis	Associate the registered users' information with the public information to estimate the health situation and make suggestions.

Table 1-3. Feature description

Besides expanding the user's information, we are also trying hardly to dig out more

creative features. Some of them are from the project description in the reference 2, and the others are thought out by ourselves. The Table 1-3 shows the features that are scheduled with fully-description. The rate of progress is shown in the plan of work section. As for the display part of our project, we choose IOS and website as our platform. The drawing APIs are PNchart[11] and Google API[12]. More details about the display part can be found in the user interface (UI) design sections.

1.2 Glossary of Terms

- Social Networking Service (SNS)

  A social networking service is a platform to build social networks or social relations among people who share interests, activities, backgrounds or real-life connections (From Wikipedia). See the Figure 1-2.

Figure 1-2. SNS

- Health Care Analytics (HCA)

  Health care analytics is a product category used in the marketing of business software and consulting services. It makes extensive use of data, statistical and qualitative analysis, explanatory and predictive modeling (From Wikipedia). See the Figure 1-3.

Figure 1-3. HCA

- Server and database

  A server is a running instance of an application (Software) capable of accepting requests from the client and giving responses accordingly. A database is an organized collection of data. The data are typically organized to model aspects of reality in a way that supports processes requiring information. For example, modeling the availability of rooms in hotels in a way that supports finding a hotel with vacancies (From Wikipedia). See the Figure 1-4.

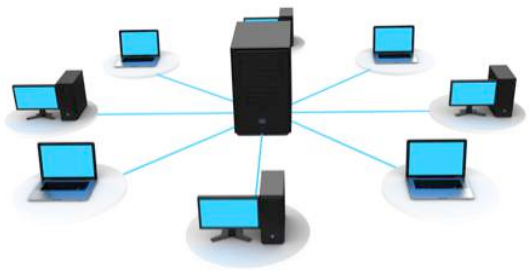


Figure 1-4. Server and database

- Application Programming Interface (API)

  In computer programming, an application programming interface (API) specifies a software component in terms of its operations, their inputs and outputs and underlying types. Its main purpose is to define a set of functionalities that are independent of their respective implementation, allowing both definition and implementation to vary without compromising each other (From Wikipedia).

- Rest API and Streaming API

  Rest API and Streaming API are both Twitter API for accessing Twitter users' data in the Twitter database. Rest API acquires historical data while Streaming API acquires real-time data. Besides, Rest API has connection limitations.

- Heat Map

  A heat map is a graphical representation of data where the individual values contained in a matrix are represented as colors (From Wikipedia). See the Figure 1-5.

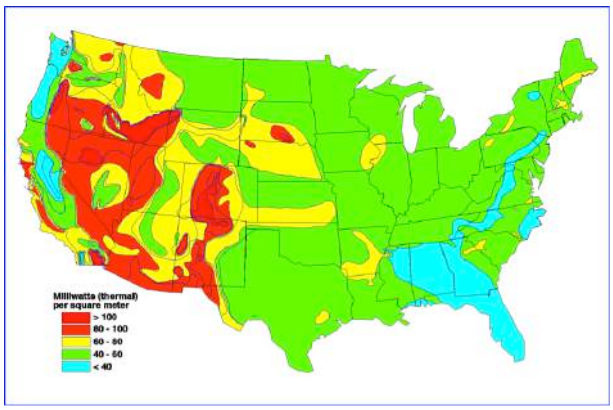


Figure 1-5. Heat map

2 System Requirements

2.1 Enumerated Functional Requirements

Identifier	PW	Requirement
REQ1a	5	The system shall attain real-time Tweets from Twitter and save them in its database.
REQ1b		The system shall acquire the historical health-related tweets the users sent before.
REQ1c		The system shall attain information of users that send health-related tweets. (Like username, location, follower numbers, Tweets Numbers, Twitter user URL, etc.)
REQ2	3	The system shall retrieve facilities' information (like location and type) from Google.
REQ3	4	The system shall acquire demographic data from demography text analytics API.
REQ4	5	The system shall obtain authorization from the third party API.
REQ5a	5	The system shall filter out unrelated tweets in the database.
REQ5b		The system shall classify relevant tweets data from the Database by different classification (like types of exercises, location, age, gender, types of food, etc.).
REQ6	4	The system shall screen out the tweets with duration of time or distance, etc.(intensity related)
REQ7	3	The system shall allocate tweets with different weight based on different types and sources.
REQ8	5	The system shall count the exercise-related numbers of tweets sent in specific area and time respectively and also calculate the percentage of exercise-related tweets.
REQ9	5	The system shall count the numbers of specific tweets and rank them. (Like running-related, swimming-related in the exercise field and like apple-related, banana-related in the diet-field, etc.)
REQ10	5	The system shall calculate the average intensity of users' exercises (including different types).
REQ11a	5	The system shall calculate the overlap between the group that post tweets about exercise and the group post tweets about wellness.
REQ11b		The system shall calculate the overlap between the group that post tweets about exercise and the group post tweets about food/diet.
REQ12	5	The system shall record same users' exercises-related tweets and calculate how regularly they do exercise.
REQ13	5	The system should show leaderboard indicating popular exercises cities, devices, topic, etc.
REQ14	5	The system shall show heat map indicating active level of certain exercise in one specific area.
REQ15	5	The system shall display the amount of different kinds of tweets (based on location, types of exercise, time, etc.) in column chart.
REQ16a	5	The system shall compare the number of exercise-related tweets in

		different time period in a given area, calculate and show the trend.
REQ16b		The system shall show the trend of amount of exercised-related tweets in a given area by displaying two Heat Maps in the same time or sequentially.
REQ17	4	The system shall display the overlap between people who concern about wellness and who exercises and also between people who exercises and who talking about diet by pie chart.
REQ18	2	The system shall display the recent related tweets on the map (based on their location) and show what they just posted.
REQ19	5	The system shall display the histogram of number of relevant tweets post by the users since they first tweet.
REQ20	5	The system shall give out a public suggestion about exercise based on the average intensity and regularity of exercise posted by twitter users.
REQ21	3	The system shall allow users to register in order to obtain personal services provided by the website.
REQ22	3	The system shall allow registered users to modify their personal profiles.
REQ23	4	The system shall allow users to search for related information of interested exercise in given cities, type of exercise and date.
REQ24	5	The system shall give out an analysis of the user's health status and propose suggestions about exercises, diet and sleep for users based on the data derived from analysis (like average intensity and regularity).
REQ25	2	The system shall recommend most active users when users search for a given exercise program or login.
REQ26	2	The system shall recommend facilities' locations when users search for a given exercise program or login.
REQ27	3	The system shall allow registered users to share the analysis result on social websites such as Twitter and Facebook.
REQ28	3	The system shall construct a score system that can reward users (by points) according to the amount and quality of their exercise.
REQ29	3	The system shall generate a leader board to rank the user based on their score.
REQ30	1	The system shall offer ways for users to share the reward. (By Twitter or Facebook)
REQ31	1	The system shall allow registered users to invite friends for new users by sending emails or post their activities in the system on Facebook or Twitter.
REQ32	1	The system shall allow advertisers to post and change their advertisements on the website.
REQ33	3	The system shall allow the administrator to modify the database.
REQ34	2	The system shall provide facilities information when the user acquires.
REQ35	3	The system shall split the words in tweets, calculate in what frequency they appear and show the result in specific forms (like in table).
REQ36	3	The system shall analyze the correlation between different health topics and show the result in some forms (like the linear regression graph).
REQ37	3	They system shall do the sentiment analysis by evaluating

		corresponding mood state the tweets show. The system shall display the results.

Table 2-1. Functional requirements

2.2 Enumerated Nonfunctional Requirements

Identifier	PW	Requirement
REQ-38	5	The system should be accessible to any users through website or mobile application (IOS).
REQ-39	4	Related information of Twitter users should be updated when changes occur.
REQ-40	4	The system should remain running if there's update in the Twitter API.
REQ-41	3	Important data should have a backup in case the system goes down.
REQ-42	3	No raw data should be dependent on third parties other than Twitter.
REQ-43	3	The system should be fixed as soon as possible every time it goes down.
REQ-44	2	The user interface for both website and application should be users-friendly and easy to navigate.
REQ-45	2	All graphs related to data collected should be displayed in a simple and direct way.
REQ-46	2	All users' information should only be stored in the database of the system.
REQ-47	2	The system should support medium or high level of testing to find faults ahead of time.

Table 2-2. Non-functional requirements

2.3 On-Screen Appearance Requirements

2.3.1 Requirement descriptions

Identifier	PW	Requirement
REQ-48a	3	Registered user clicks on "Sign In" button to log in our application.
REQ-48b	3	People who want to register can click on the "Sign Up" button. They will be lead to a sign up window.
REQ-48c	4	This window shows recent tweets which about exercise.
REQ-48d	4	It offers both Android and IOS app to users in order to get a better user experience.

REQ-48e	3	User can share our website with other SNS.
REQ-49a	4	It allows the user to search for cities, communities, neighbor-hoods, relevant hash- tags, or other users.
REQ-49b	4	User scan find their partner by searching location, exercise program and time.
REQ-49c	5	Add leaderboard with more classifications that tweet about health in different states and city.
REQ-49d	5	This window shows data analysis about the amount of people in a given area exercise and the timeline of people exercise, etc.
REQ-49e	5	This window shows the frequency of exercise by different states and cities. Meanwhile, any new tweet can be found on this map.
REQ-50	3	On this page, you can find a lot of information about your entered community, like Health score, Exercise rank, Total number of twitter exercisers, Exercisers percentage, Exercise frequency, Exerciser distribution, Popular sports, devices and gyms.
REQ-51	4	On this page, you can view device introduction, positive experience, negative experience, device overall grade, and comments. Clicking on the “share” button, user will be linked to a “share your comment” page. Clicking on the name in the comments module, user will be linked to this reviewer’s twitter home page.
REQ-52	3	This page is where the user manages their personal account such as changing their login or password or location, etc.

Table 2-3. On-screen appearance requirements

2.3.2 Website appearance

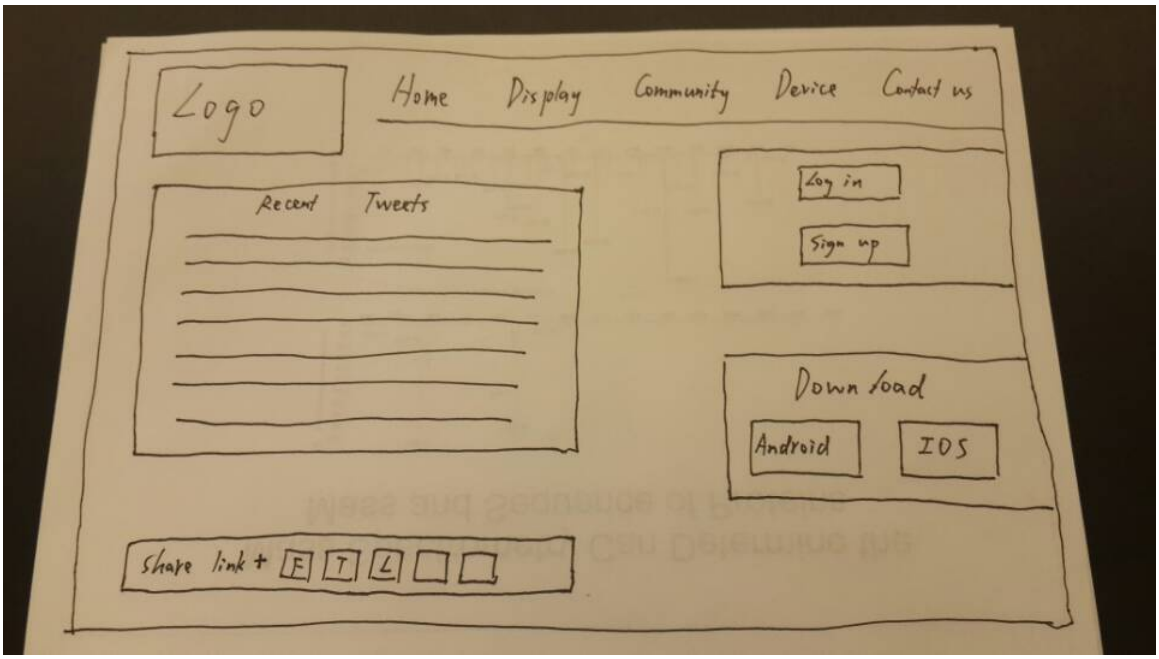


Figure 2-1. Homepage of the website

The figure 2-1 shows the homepage of our website. The users can easily see the menu: Home, Display, Community, Device and Contact Us. When clicking the Home menu, the users can see the recent tweets about health and exercise. Also, members can login to manage their personal profile. The Android and IOS apps are provided to free download. When clicking the Display menu, the users can browse several analysis charts or graphs about health and exercise. Details are shown in figure 2-2. When clicking the Community menu, members who has logged in can use the forum to contact with other members by asking or answering questions. When clicking the Device menu, the users can see the leaderboard of device recommended by members. When clicking the Contact us menu, the users can contact us by making any suggestion or improvement.

Figure 2-2. Display page

The figure 2-2 shows the details about the display menu. Firstly, the users can use the search bar to search exercise they like. Secondly, the leaderboard shows the rankings of different exercise such as running and swimming. Thirdly, the users can see the analysis of tweets about health and exercise, in order to know the amount of people divided by area, time, demographics and exercise type. Also, the state map and marker map are provided to browse.

2.3.3 IOS application appearance

Figure 2-3. Main menu

In the Figure 2-3, this interface contains two list views, “Public Information” and “User Profile”. Each item in the “Public Information” list shows the specific analysis of the public who are tweeter users. Each item in “User Profile” list shows analysis concerning user’s own tweet data.

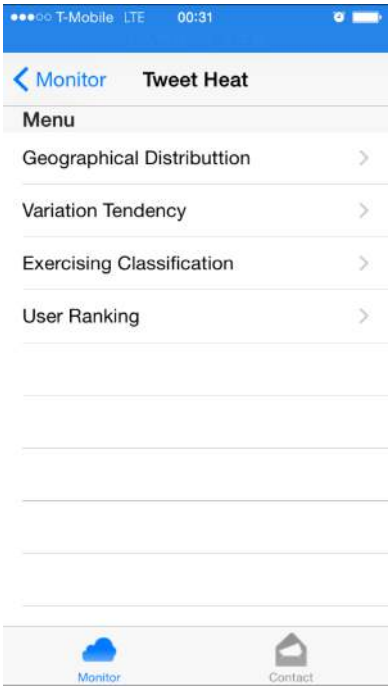
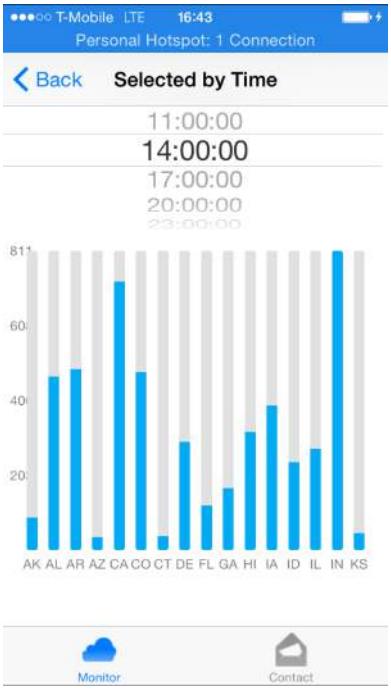


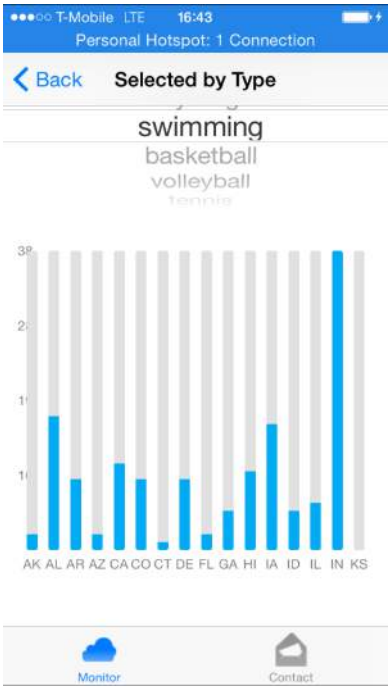
Figure 2-4. Tweet heat menu

In the Figure 2-4, this interface contains a sub menu in which each item contains analytical information concerning the exercise intensity, distribution of the public.



In the Figure 2-5, this interface shows the bar chart of the exercise intensity of different states in a certain time section selected by the user.

Figure 2-5. Geographical distribution (a)

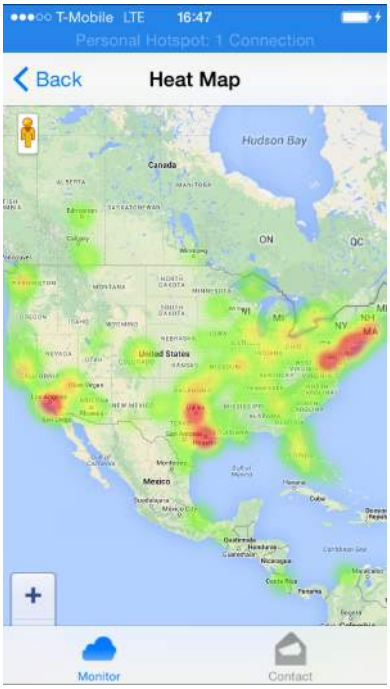


In the Figure 2-6, this interface shows the bar chart of the exercise intensity of different states in a certain exercise type selected by the user.

Figure 2-6. Geographical distribution (b)

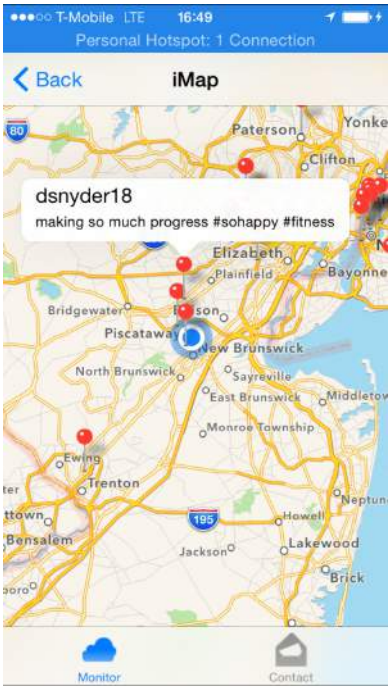
In the Figure 2-7, this interface shows a US map in which each state contains information of the amount of tweets concerning exercise and health. The more the number of tweets being tweeted, the darker the color of the certain state will be.

Figure 2-7. Geographical distribution (c)



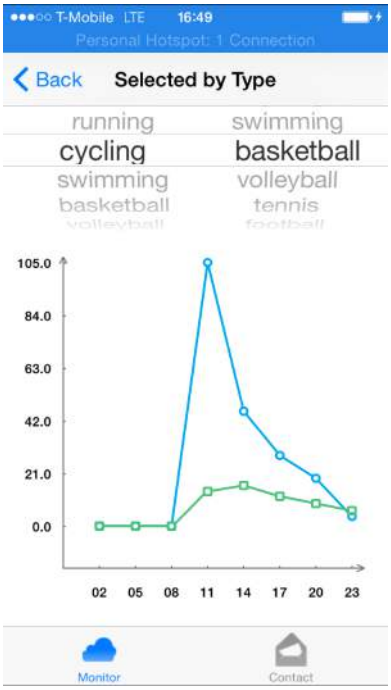
In the Figure 2-8, this interface shows the exercise intensity in different areas all over the world. The red spot shows that there are more exercise lovers, while the green spot shows there are less exercise lovers.

Figure 2-8. Geographical distribution (d)



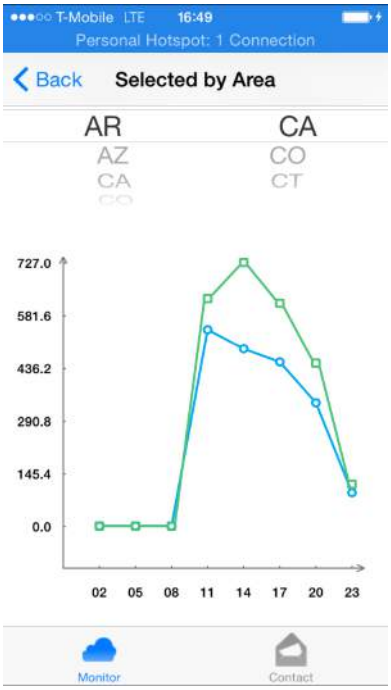
In the Figure 2-9, this interface shows a Google map where marks will be showed on it if users post their tweet with location info. The interface will pop out certain user's name and his tweets if a single mark is clicked.

Figure 2-9. Geographical distribution (e)



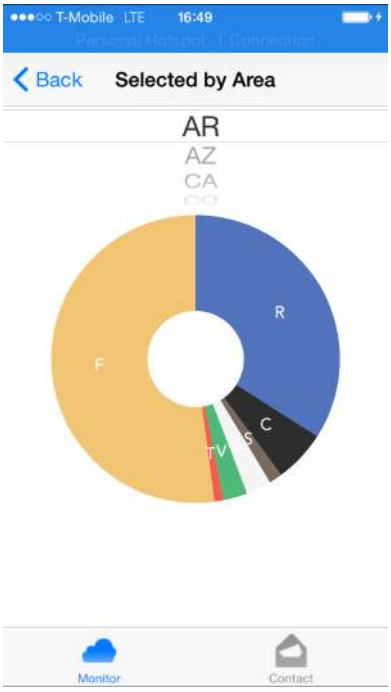
In the Figure 2-10, This interface shows the line chart of the exercise tendency in different exercise types as time goes by. Users are allowed to choose two different types at the same time to compare their tendencies.

Figure 2-10. Variation tendency (a)



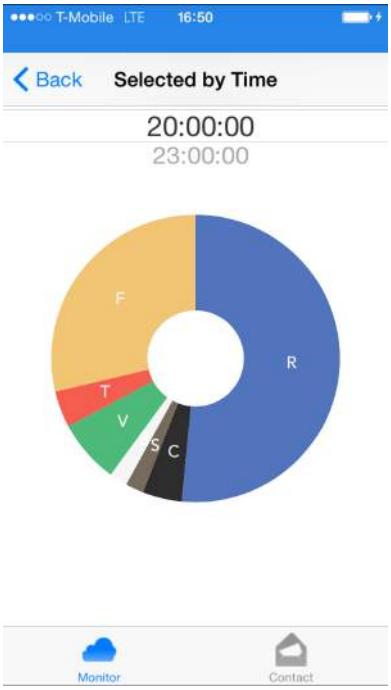
In the Figure 2-11, this interface shows the line chart of the exercise tendency in different areas as time goes by. Users are allowed to choose two different states at the same time to compare their tendencies.

Figure 2-11. Variation tendency (b)



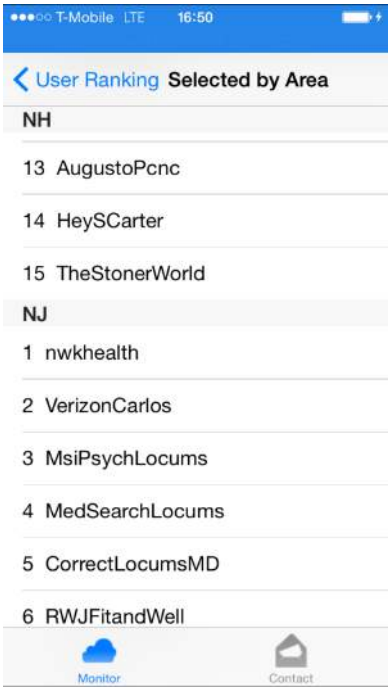
In the Figure 2-12, this interface shows the pie chart of the popularity of different exercise types as in a certain state chosen by user. Each letter represents a certain exercise type (R-running, C-cycling, S-swimming, B-basketball, V-volleyball, T-tennis, F-football).

Figure 2-12. Exercising classification (a)



In the Figure 2-13, this interface shows the pie chart of the popularity of different exercise types as in a certain time section chosen by user. Each letter represents a certain exercise type (R-running, C-cycling, S-swimming, B-basketball, V-volleyball, T-tennis, F-football).

Figure 2-13. Exercising classification (b)



In the Figure 2-14, this interface displays the ranking list of the users who exercise the most often in different states.

Figure 2-14. User ranking (a)

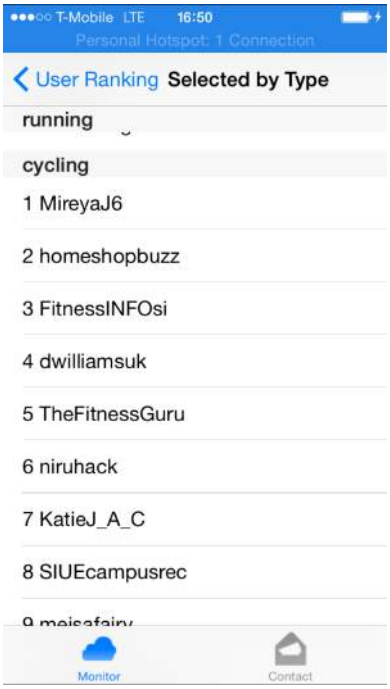
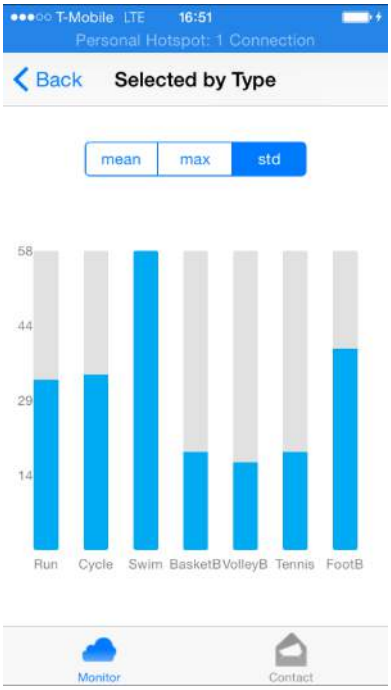


Figure 2-15. User ranking (b)

In the Figure 2-15, this interface displays the ranking list of the users who exercise the most often in different exercise types.

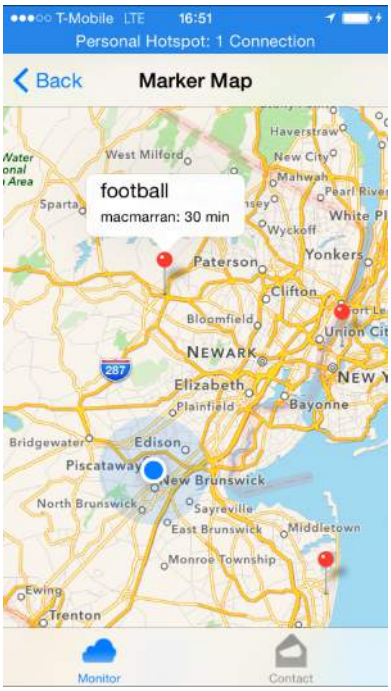
Figure 2-16. Exercising duration (a)

In the Figure 2-16, this interface shows the analysis of time duration of people's exercising in different areas, providing "average", "maximum", and "individual difference" of people's exercising duration time.



In the Figure 2-17, this interface displays a bar chart of the "mean", "max" and "standard deviation" of exercise duration time varied by different exercise types.

Figure 2-17. Exercising duration (b)



In the Figure 2-18, this interface shows markers on the map for users who have tweeted a tweet concerning exercise and mentioned his exercise duration time. Once the marker is clicked, the interface will pop out the duration time and the exercise type on the top of the marker.

Figure 2-18. Exercising duration (c)

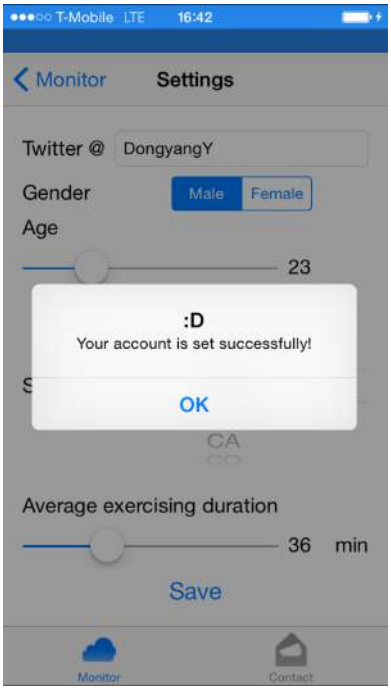


Figure 2-19. Login

In the Figure 2-19, in this interface the user can login to load their personal profile by inputting Twitter screen name, gender, age, location and average exercising duration.

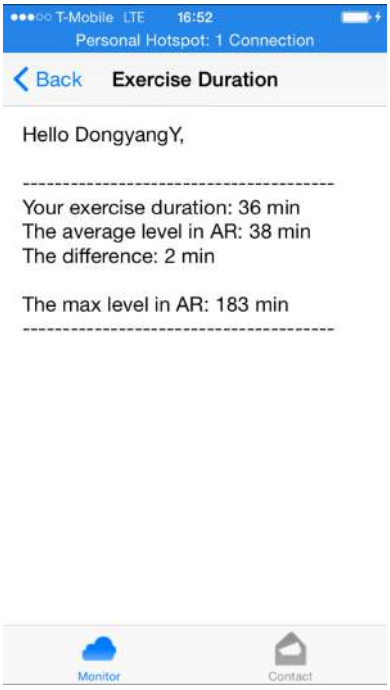


Figure 2-20. Personal diagnosis

In the Figure 2-20, in this interface the users can see an analysis of their personal exercise history. Personal exercise duration, average level in area and the difference can be shown.

3 Functional Requirements Specification

3.1 Stakeholders

Our system focuses on offering exercise and health field problems solutions, so the stakeholders include these people and organizes:

- System Users

  Defined as most important customers. For exercise topic, they can use our system to view exercise heat map, leaderboard, score system, to find friends and gym location, to find suitable sports for individual. For healthy related topic they can find useful information about diet and sleep on twitter via our system, can acquire suggestions on health issues through our system.

- Systems architects and developers

  Defined as system supporters. They take responsibility on building the system that fulfills the user's requirements, testing and maintaining the system, and provide technical support to other stakeholders.

- Academic Researchers

  Defined as third party, can use our system to gain information about people's health and exercise situation for academic research purpose.

- Gym and sports equipment advertisers

  Defined as our profit supporters, can pay to put advertisement in our system. Advertisement is restricted to healthy and exercise fields.

3.2 Actors and Goals

- User (Initiating type)

Goals: to interact with the system, acquire exercise and health information they need, make friends and find useful exercise places via the system.

- Administrator (Initiating type)

  Goals: has the top priority to collect data, access, manage, and maintain the database, provide service to the user.

- Advertiser: (Initiating type)

  Goals: analyze data, put and manage advertisement on the system to attract customers to buy their commodities.

- Google server and database (Participating type)

- Twitter server and database (Participating type)

- Demography text analytics server and database (Participating type)

- Our server and database (Participating type)

3.3 Use Cases

3.3.1 Casual description

The summary use cases are as follow:

- UC2 User information adding

  Allow the user to create account in the Health Monitoring System for some private health-related services.

  Derived from REQ20, 21.

- UC3 Advertiser information adding

  Allow the administrator to create account in the Health Monitoring System for advertisers.

  Derived from REQ20, 21.

- UC4 Data deleting

  Allow the administrator to delete useless, incorrect data or do some necessary adjust in system's database.

  Derived from REQ32.

- UC5 Data collecting & classifying

  Allow the administrator to retrieve Twitter's users' data from Twitter and demography text analytics database, classify them by some specific sort (like location, exercise type, etc.) and store these data in system's database.

  Derived from REQ1-3, 5-7.

- UC6 Third party API auth.

  Allow the administrator to get verification of accessing and retrieving data from Twitter and demography text analytics.

  Derived from REQ4.

- UC8 Advertisement updating

  Allow the advertiser to change the advertisement post on the website after login.

  Derived from REQ31.

- UC11 Exercise heat

  Allow the public user to have a glance at exercise heat in area, time, type, demography and also their trends. (Shown by heatmap, pie chart, column chart, etc.)

  Derived from REQ14-16.

- UC12 Public exercise suggestions

  Allow the public user to acquire suggestions about proper frequency and intensity of exercise. The suggestion varied by different demography and types.

  Derived from REQ20.

- UC13 Public ranking

Allow the public user to see the ranking of different area, exercise types and demography by amount of tweets.

Derived from REQ13.

- UC14 Correlation between health topics

  Allow the public user to check the overlap between people who concern about wellness and who exercises and also between people who exercises and who talking about diet by pie chart.

  Derive from REQ17.

- UC15 Facilities search

  Allow the public search facilities in a given area he/she chooses. Show detail information about the facilities include their location and main services. Derived from REQ34.

- UC16 Score system

  Allow the login user to see about their overall score and their ranking in corresponding area and demography (The score is allocated based on the exercise-related tweets he/she posted before).

  Derived from REQ28-30.

- UC17 Personal suggestions

  Allow the login user acquires suggestions about how their exercise should be (including intensity and regularity), nearest facilities, recommendation of friends (like twitter users that share the same hobby and have high activities).

  Derived from REQ24-27.

- UC18 User's record

  Allow the user to have a look at his/her own exercises history on Twitter. The record is shown by histogram, including exercise time and types.

  Derived from REQ19.

- UC19 Login

Allow the user (including the normal user and advertiser) to login and gain specific services (Like normal user would have the system analyzed their health status and provided he/she related services, advertiser could login to change their previous advertisement).

Derived from REQ21.

- UC20 Data analyzing

  Allow the advertiser to analyzing the data stored in database like count the total and exercise-related numbers of tweets, rank different types tweets by amount, calculate the average intensity of users' exercises (including different types), etc.

  Derived from REQ8-12.

3.3.2 Use case diagram

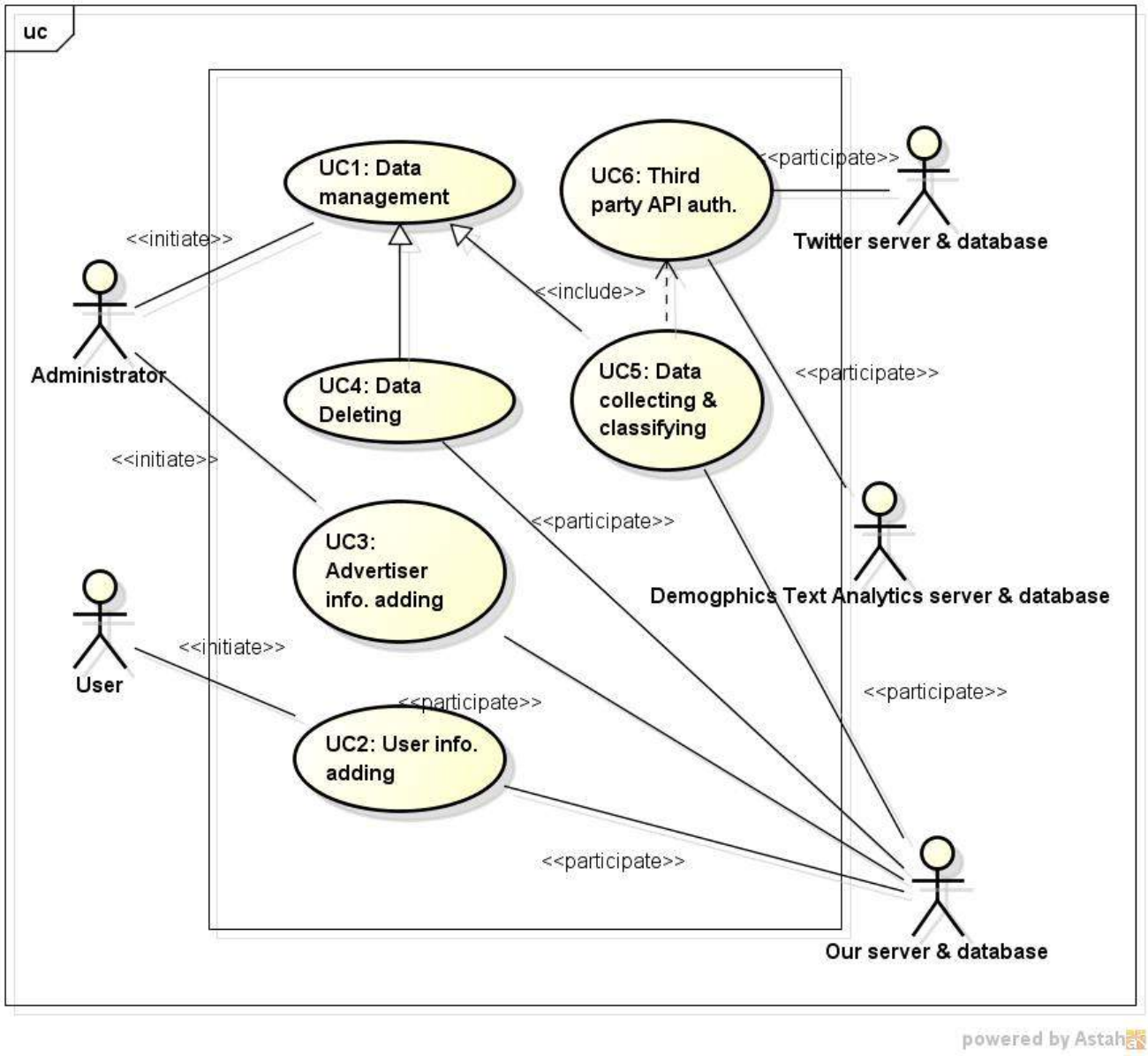


Figure 3-1 Use case diagram of server and database management subsystem

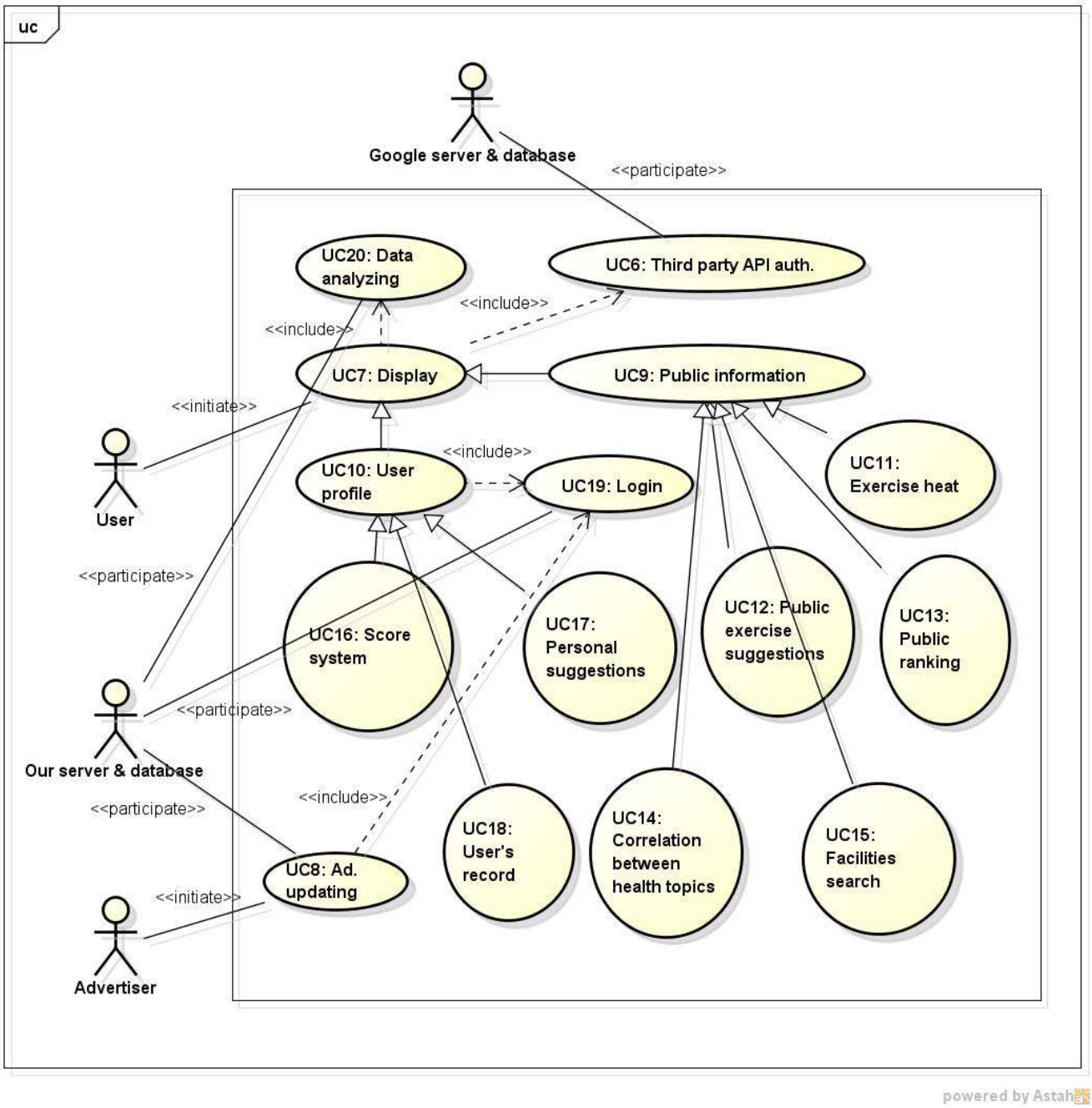


Figure 3-2 Use case diagram of displaying subsystem

3.3.3 Traceability matrix

REQ	PW	UC 2	UC 3	UC 4	UC 5	UC 6	UC 8	UC 11	UC 12	UC 13	UC 14	UC 15	UC 16	UC 17	UC 18	UC 19	UC 20
1a	5				√												
1b					√												
1c					√												
2	3				√												
3	4				√												
4	5					√											
5a	5				√												
5b					√												
6	4				√												
7	3				√												
8	5																√
9	5																√
10	5																√
11a	5																√
11b																	√
12	5																√
13	5									√							
14	5							√									
15	5							√									
16a	5							√									
16b								√									
17	4										√						

18	2															
19	5													√		
20	5								√							
21	3	√	√												√	
22	3	√	√													
23	4							√								
24	5												√			
25	2												√			
26	2												√			
27	3												√			
28	3											√				
29	3											√				
30	1											√				
31	1															
32	1						√									
33	3			√												
34	2										√					
35	3									√						√
36	3									√						√
37	3								√	√			√			√
Max PW		3	3	3	5	5	1	5	5	5	4	2	3	5	5	3	5
Total PW		6	6	3	24	5	1	19	8	8	10	2	7	15	5	3	34

Table 3-1 Traceability matrix

3.3.3 Fully-dressed description

Use Case UC-5:	**Data collecting & classifying**
Related Requirements:	REQ1, REQ2, REQ3, REQ5, REQ6, REQ7
Initiating Actor:	Developer
Actor's Goal:	To find useful information using keyword and store them in our database
Participating Actors:	Twitter's server & database, System's server & database
Pre-conditions:	·Developer has been authorized to crawl information from Twitter ·Database is all set and ready to store new data
Post-conditions:	Useful information from related Tweets is stored in system's database
Flow of Events for Main Success Scenario: → 1. The developer sets up the database and decide the keywords for using Twitter's API ← 2. Twitter gives out the information filtered by the keywords → 3. The developer collects the data and store them in the system's database	

Table 3-2 Fully-dressed description UC-5

Use Case UC-11:	**Exercise heat**
Related Requirements:	REQ14, REQ15, REQ16
Initiating Actor:	User
Actor's Goal:	To get visual charts or diagrams of the information offered by the system
Participating Actors:	System's server and database, Google server & database
Pre-conditions:	Related information is stored in the system's database
Post-conditions:	System displays the chosen diagrams on the screen
Flow of Events for Main Success Scenario: → 1. The user navigate to the exercise intensity interface → 2. The user choose certain distribution and type to see the statistical graphs ← 3. The system displays the diagrams as requested	

Table 3-3 Fully-dressed description UC-11

Use Case UC-15:	**Facilities search**
Related Requirements:	REQ34
Initiating Actor:	User
Actor's Goal:	To search the nearby exercise facilities on Google map
Participating Actors:	Google server & database
Pre-conditions:	·System has the authorization to get information of user's location ·System has the access to Google API

Post-conditions:	The location and some other information of the nearby exercise facilities are showed on the Google map
Flow of Events for Main Success Scenario: → 1. The user navigate to the facility search interface → 2. The system gets the current location of the user → 3. The user gives the exercise facility type ← 4. The system displays all nearby facilities of certain type on the Google map	
Flow of Events for Extension (Alternate Scenario): → 1. The user navigate to the facility search interface → 2. The user choose a certain area → 3. The user gives the exercise facility type ← 4. The system displays facilities of certain type in the chosen area on the Google map	

Table 3-4 Fully-dressed description UC-15

Use Case UC-17:	**Personal suggestions**
Related Requirements:	REQ24, REQ25, REQ26, REQ27
Initiating Actor:	User
Actor's Goal:	To acquire suggestions about their daily exercise, facilities, devices, and recommendations of partners with similar interests
Participating Actors:	System's server & database, Twitter's server & database
Pre-conditions:	·User has logged in onto the system ·System has the authorization to get information of user's profile
Post-conditions:	Suggestions are showed on the website in the form of useful links
Flow of Events for Main Success Scenario: → 1. The user navigate to the main interface of the system → 2. The user logs into the system using their username and password ← 3. The system lists out the related links concerning their daily exercise and the profile links of Twitter users with same interest	

Table 3-5 Fully-dressed description UC-17

Use Case UC-20:	**Data analyzing**
Related Requirements:	REQ8, REQ9, REQ10, REQ11a, REQ11b, REQ12
Initiating Actor:	Developer
Actor's Goal:	To obtain the statistical information from the data collected from Twitter
Participating Actors:	System's server & database
Pre-conditions:	·Related data are stored in the system's database ·Collected data are assumed to be correct
Post-conditions:	All statistical data are stored in different tables in the

	database

Flow of Events for Main Success Scenario:

→ 1. The developer chooses certain data from the database

→ 2. The developer decides methods to analyze the selected data

← 3. The database gives out the statistical result of the analysis

Table 3-6 Fully-dressed description UC-20

3.3 System Sequence Diagrams

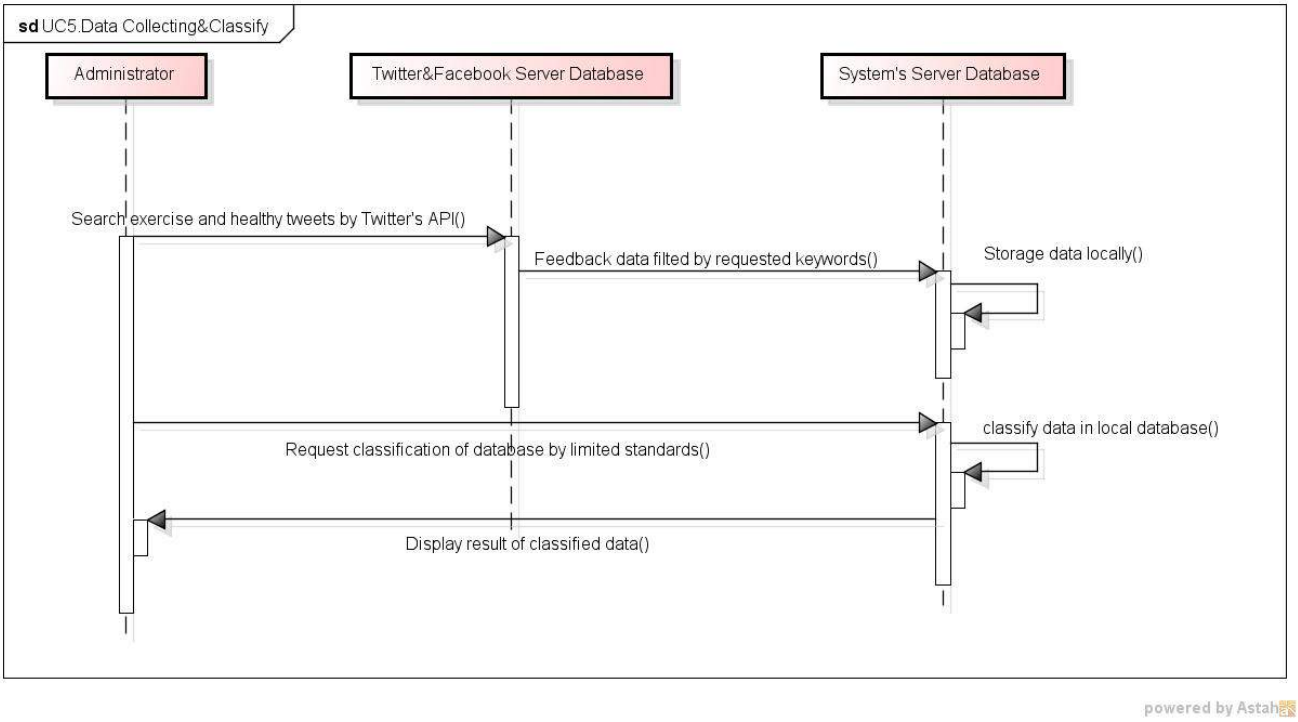


Figure 3-3 Use case diagram for UC-5

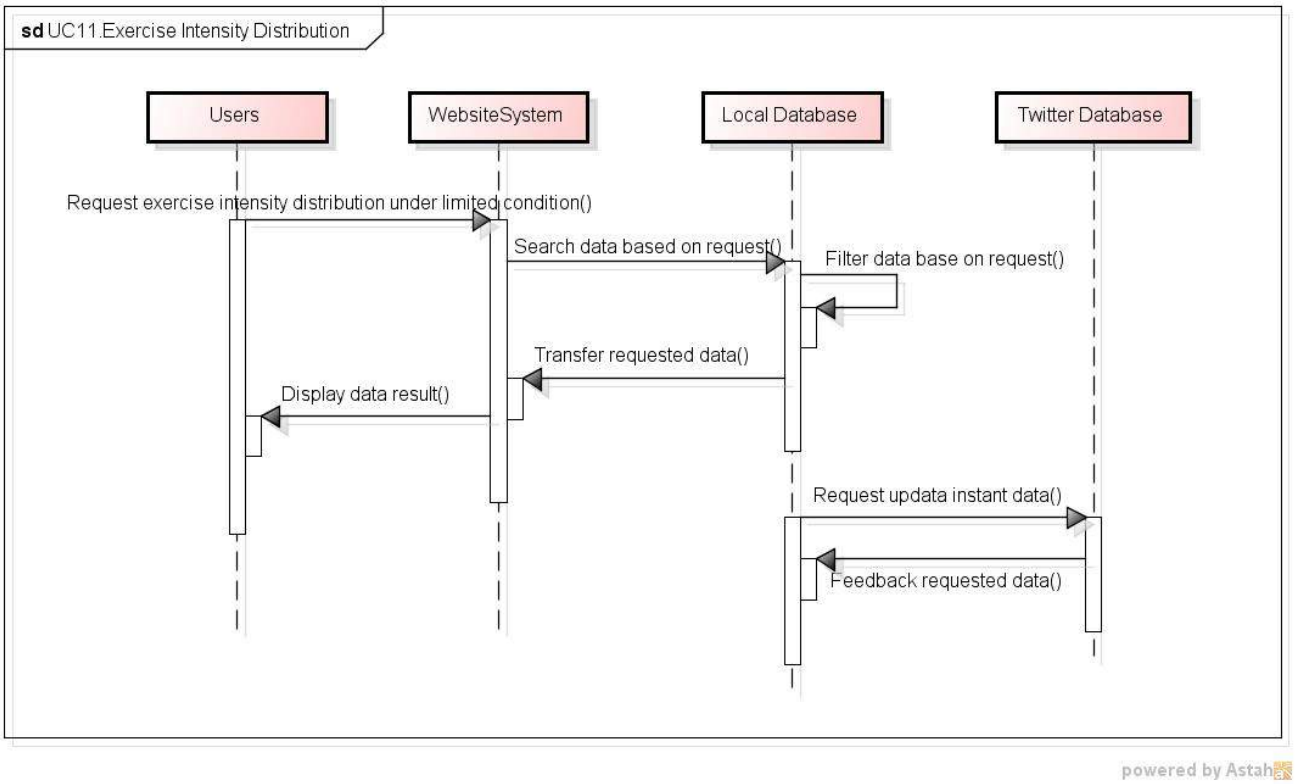


Figure 3-4 Use case diagram for UC-11

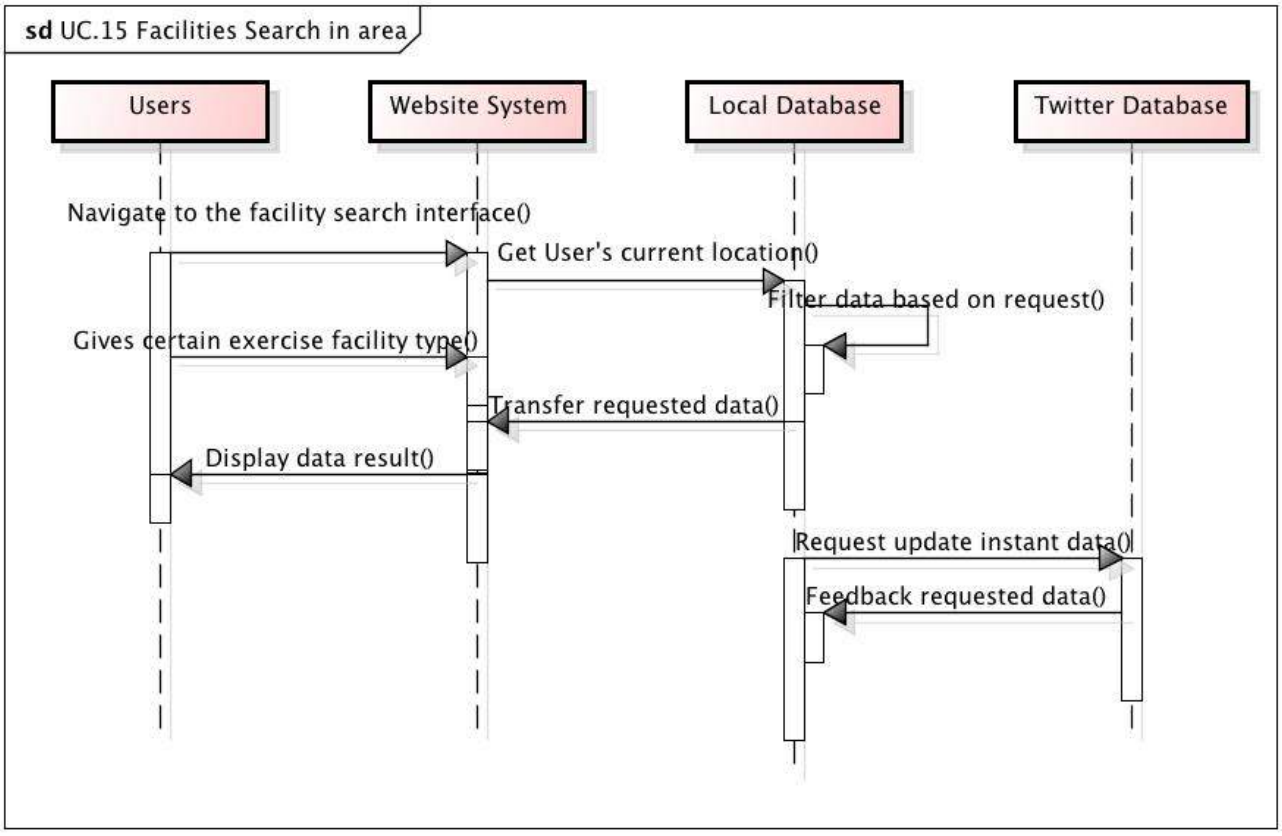


Figure 3-5 Use case diagram for UC-15

4 User Interface Specification

4.1 Preliminary Design

This section represents our preliminary design and analysis. Here is the proposed main user interface webpage.

- The visitor (user who has not registered) can browse both “Display” and “Search” sections. In “Display” section, the visitor can see the leaderboard by selecting different buttons, which includes ranking in users varied by area, type and demography.

- The visitor is able to see different heat maps varied by area, time, type and demography. He/she can select the display way such as in area, time, type or demography. The map also enables dragging and zooming in/out.

- The visitor is able to see different state maps varied by time, type and demography.

- The visitor is able to see marker maps, which can show the real-time update whenever someone post a new tweet about health or exercise.

- If the visitor wants to know the correlation/overlap between the groups that exercises and the group that discusses health and wellness, he/she can easily fins the result by browsing the part of “correlations”.

- The visitor may get some suggestion about frequency, time and amount of exercise varied by demography and exercise type.

- In “Search” section, the visitor can search facilities in area and find partners who share the same exercise type and time.

- The member (user who has registered) can use these features stated above that all visitors can use.

- The member can log in to use some additional features about personal profile, such as personal overall ranking, exercise record and suggestion.

- After logging in, the member is able to know his/her overall ranking in all of our website members. The ranking is varied by exercise type.

- The member can also check personal record of exercising in time varied by type.

- The website offers particular suggestion of exercising, facility, device and friend, helping the member enjoy the process of exercising in a better way.

Above are the main features for our website. More features would be added in as stated in the system requirements.

4.2 User Effort Estimation

Our website is very easy to use. We try to design it with the minimum user effort to accomplish their goal.

For the visitor who just wants to check the health activities awareness in certain city and obtain some statistical data:

- NAVIGATION (several keystrokes and one click)

  Navigate to our software webpage (several keystrokes; inputting http address)
  Main interface page is brought to the visitor
  Close our webpage when finished (one click)

- DATA ENTRY (several keystrokes and clicks)

  Select one button from “area”, ”type” or “demography” to see different rankings (one click)
  Select one kind of heat map distribution (one click)
  See the correlation/overlap between the group that exercises and the group that discusses health and wellness (one click)
  Select one button from “area”, ”type” or “demography” to see exercise suggestions (one click)
  Select location and exercise type to search facilities nearby. (three clicks)
  Select location, time and exercise type to find partners. (four clicks)
  Select one kind of heat map distribution (one click)

For the visitor who wants to register to a member:

- REGISTRATION NAVIGATION (two clicks)

  Click on log in button (one click)
  A new page pops up asking for user name and password, and an option of

registration. (0 effort)
Click on the register link (one click)
A registration page pops up asking for information
Done with registration.

- INFORMATION FILLING (several keystrokes)

  Account Registration Part 1 (Instructions and how to use the application)
  Account Registration Part 2 (Disclaimers and Permissions)
  Account Registration Part 3 (User information)
  Done with registration and a personal page is set at the same time.

For the member who wants to log in:

- LOG IN NAVIGATION (one click)

  Click on log in button (one click)
  A new page pops up asking for user name and password. (two keystrokes)
  Done with registration.

5 Domain Analysis

5.1 Domain Model

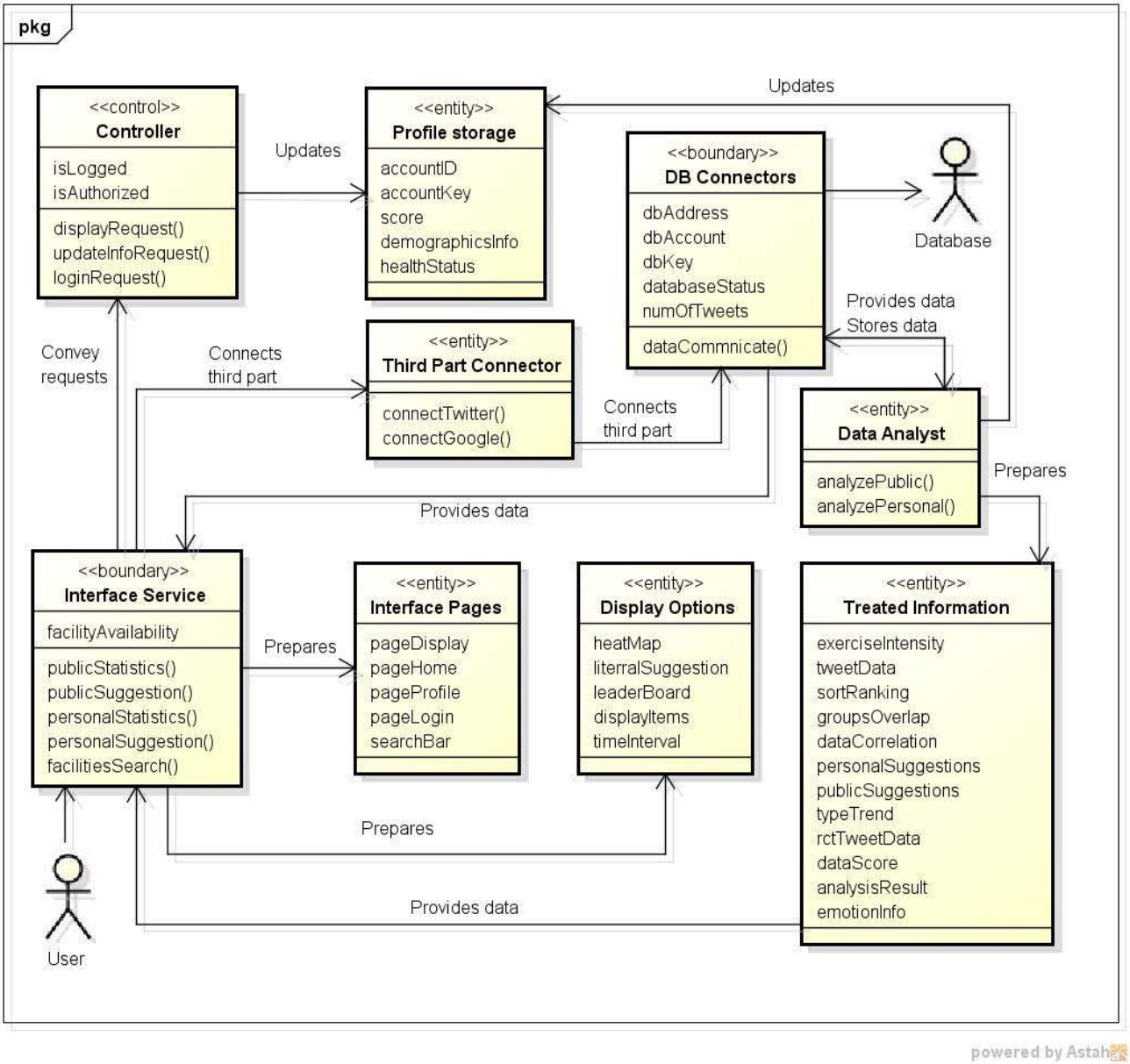


Figure 5-1 Domain model diagram

5.1.1 Concept definitions

We derive the domain model concepts from detailed user cases. Table 5-1 lists the

responsibilities and the assigned concepts.

First, the responsibilities 1, 2, 5, 6, 7, 8, 9, 10, 11, 14 are basically identified from the scenarios of UC11, 12, 13, 14, 15. All of them are related to the information display and the interaction between users and the system. Second, the responsibilities 3, 4 are derived from UC2, 3, 8, 16, 17, 18, 19. They deal with the login issues and personal service provided by the system. At last, the responsibilities 12, 13 generate from UC5, 20. Both of them are components that assist the system to analyze and classify tweets data.

Responsibility	Type	Concept
R1: Prompt the user to make movement for available services.	D	Controller
R2: Handle requests from users.	D	Controller
R3: Deal with login issue (check users key, approve or deny).	D	Controller
R4: Container for the collection of valid keys and account profile associated with users.	K	Profile Storage
R5: Show pages for user to create account, login and logout.	D	Interface Service
R6: Show search engine and diversity display buttons for user to choose.	D	Interface Service
R7: Display related information in literal, numerical, graphical and map forms.	D	Interface Service
R8: Static websites and mobile phone interface that shows the user the current context, what services could be used, and outcomes of the previous request.	K	Interface Pages
R9: Specific parameters and options for information display, including search options, types of graph and display items.	K	Display Options
R10: Related information after all analysis, inferring and statistics for display.	K	Treated Information
R11: Manage interactions with the database.	D	DB Connector
R12: Classify, do statistics, analyze and infer related information for suggestion, statistics display and recommendation.	D	Data Analyst
R13: Retrieve related data from Twitter Database.	D	Third Part Connector
R14: Retrieve related services from Google Database.	D	Third Part Connector

Table 5-1 Concept definitions

5.1.2 Association definitions

Some concepts defined above need to work together in order to achieve specific

functions. The concepts work together are called concept pair. The definition and description of concept pair of the system are listed in Table 5-2.

Concept pair	Association description	Association name
Controller ↔ Profile Storage	Controller updates User Profile when the user change his/her information.	Updates
Controller ↔ Interface Service	Controller passes requests to Interface Service and receives pages prepared for displaying.	Convey requests
Interface Service ↔ Interface Pages	Interface Service prepares the Interface Pages.	Prepares
Interface Service ↔ Display Options	Interface Service prepares the Display Options.	Prepares
Interface Service ↔ DB Connector	Database Connection passes the retrieved data to Interface Service to render them for display and show.	Provides data
Interface Service ↔ Treated Information	Interface Service extracts information from Treated Information for display.	Provides data
Interface Service ↔ Third Part Connector	Third Part Connector enable Interface Service to connect the Third Part to ask for service (Like the google map and graphs).	Connects Third Part
DB Connector ↔ Third Part Connector	Third Part Connector enable DB Connector to connect the Third Part to retrieve related data.	Connects Third Part
DB Connector ↔ Data Analyst	Database Connection passes the retrieved data to Data Analyst for analysis, inferring and statistics. Data Analyst stores useful data back into Database after analysis.	Provides data, Stores data
Data Analyst ↔ Treated Information	Data Analyst prepares the Treated Information.	Prepares
Data Analyst ↔ Profile Storage	Data Analyst changes the information (like score and health status) in Profile Storage.	Updates

Table 5-2 Association definitions

5.1.3 Attribute definitions

Attributes of domain concepts are derived in Table 5-3.

Concept	Attribute	Attribute Description
Controller	displayRequest	Send user's requests to retrieve display service
	updateInfoRequest	Send user's requests to retrieve profile change service
	loginRequest	Use the data the user input to request a login operation
	isLogged	Identity parameter to determine whether the user is login
	isAuthorized	Identity parameter to determine whether the user is authorized
Profile Storage	accountID	Identity number used to determine the user
	accountKey	Specific key to determine the user's credentials
	score	User's current score in the score system
	demographicsInfo	User's personal information like age, gender, location, etc.
	healthStatus	User's exercise-related status given by the analyst
Interface Service	publicStatistics	Show the information about public health-related statistic
	publicSuggestion	Show the exercise suggestions for the whole public
	personalStatistics	Show the information about personal health-related statistic
	personalSuggestion	Show the health-related suggestions for user
	facilitiesSearch	Search for facilities information with given conditions
	facilityAvailability	Identity parameter to determine whether facility is available
Interface Pages	pageDisplay	Pages for related information display.
	pageHome	Home page of the website or application.
	pageProfile	User's profile information page.
	pageLogin	Section on the website or in the application for user to login
	searchBar	In site search bar for user to search for related display service
Display Options	heatMap	Heat map for showing intensity or amount of relation data
	graph	Graph type information display
	literalSuggestion	Literal suggestion given out by the analyst showed in word
	leaderBoard	Leaderboard used to show the ranking of different data
	displayItems	Specific items for user to choose for display, such items are set as exercise amount, intensity, cites, etc.
	timeInterval	Time interval for user to choose for information display
Treated Information	exerciseIntensity	People's exercise regularity and intensity calculated by analyst

	tweetData	Data of amount and location of health-related tweets
	sortRanking	Amount, intensity or other attributes of tweets ranking among different set like cities, states, users, exercise types, etc.
	groupsOverlap	Data of the overlap between people who concern about wellness and who exercises and also between people who exercises and who talking about diet
	dataCorrelation	Correlation between different type of health-related tweets
	personalSuggestions	Personal suggestion include exercise intensity and regularity suggestion, personal ranking, etc.
	publicSuggestions	Public suggestion based on the average intensity and regularity of exercise among the whole country
	typeTrend	Trend of amount, intensity, regularity in different exercise type
	rctTweetData	Contents, location, user information of related tweets collected recently
	dataScore	User's data in score system, including exercise score, ranking in the system, award based on the score, etc.
	analysisResult	Analysis based on User's historical exercise data, like average exercise intensity and regularity, etc.
	emotionInfo	Data about distribution, amount and type of emotion extracted from tweets.
DB Connector	dbAddress	Address of relation database
	dbAccount	Account to manage the database
	dbKey	Key to manage the database
	databaseStatus	Identity parameter to determine whether the database is open
	numOfTweets	total number of parsed tweets
	dataCommunicate	Retrieve and store data from third part server, due with data communication inside system.
Data Analyst	analyzePublic	Analysis for public information, including statistics data, suggestion inferring, trends calculating, correlation analysis, etc.
	analyzePersonal	Analysis for personal information, including historical record, personal health status deduction, specific suggestions, score calculating, etc.
Third Part Connector	connectTwitter	Connect Twitter by API Auth. to retrieve tweets
	connectGoogle	Connect Google by API Auth. to retrieve map, chart and place finding services

Table 5-3 Attribute definitions

5.1.4 Traceability matrix

Table 5-4 shows how the system use cases map to the domain concepts. It is generated according to the responsibilities of concepts defined above.

		Domain Concepts								
Use Case	PW	Controller	Profile Storage	Interface Services	Interface Pages	Display Options	Treated Information	DB Connector	Data Analyst	Third Part Connector
UC2	6	√	√	√	√			√		
UC3	6	√	√	√	√			√		
UC4	3							√		
UC5	24							√	√	
UC6	5							√		√
UC8	1	√	√	√	√			√		
UC11	19			√	√	√	√	√		√
UC12	5			√	√		√	√		
UC13	5			√	√	√	√	√		√
UC14	4			√	√	√	√	√		√
UC15	2			√	√	√	√	√		√
UC16	7		√	√	√			√		
UC17	12		√	√	√		√	√		
UC18	5		√	√	√		√	√		
UC19	3	√	√	√	√			√		
UC20	25						√	√	√	
Max PW		6	12	19	19	19	19	25	25	19
Total PW		16	40	75	75	30	52	132	49	35

Table 5-4 Traceability matrix

5.2 System Operation Contracts

Operation	Data collecting & classifying
Preconditions	·Developer get authorized access to Twitter API through OAuth ·databaseStatus = “open” ·numOfTweets = 0, for the initialization of the database
Postconditions	·databaseStatus = “closed” ·All JASON data are parsed into the database of the system ·numOfTweets = total number of parsed tweets

Table 5-5 Data collecting & classifying

Operation	Exercise heat
Preconditions	·Related information all gets analyzed and stored in the database ·Developer get authorized access to Google API
Postconditions	System displays the Google heat map to show the intensity of people who exercise regularly

Table 5-6 Exercise heat

Operation	Facilities search
Preconditions	·System has the authorization to get information of user's location ·Developer get authorized access to Google API ·facilityAvailabiliy = true, for all facilities nearby to be displayed
Postconditions	The location and some other information of the nearby exercise facilities are showed on the Google map

Table 5-7 Facilities search

Operation	Personal suggestions
Preconditions	·isLogged = true, and ·isAuthorized = true, for current users who has logged in to our system and approved the authorization
Postconditions	·isLogged and isAuthorized remain unchanged ·Suggestions are showed on the website in the form of useful links

Table 5-8 Personal suggestions

Operation	Data analyzing
Preconditions	·Related data are stored in the system's database ·System's database is able to be changed
Postconditions	·New tables have been created into the database ·Statistical data are stored in different tables

Table 5-9 Data analyzing

5.3 Mathematical Model

5.3.1 Improve data reliability using weight index

When we set key words and collect data, there is unavoidable noise in the result, some tweet text may reflect our searching intention, some may have no relationship with the health topic. For example, if we use run as key words, the result tweet text may contain runny or rune. Thus when we do data analyze, we need to eliminate irrelevant tweet text and keep useful information. The methods we use is putting weight index to each kind of sport. And the methods will implement in emotion analyze and heat map analyze feature.

1. Weight index math model description

The basic weight index equation shows below:

If weight index is greater than 80%, we define the keyword as reliable, otherwise we define the keyword as unreliable.

However, because the database is too large to test them all, so we use systematic sample method to get limited tweet text for each key words. The above equation will change to

$$\textit{key}\text{ word weight index} = \frac{\textit{each}\text{ field }\textit{sample}\text{ veritable tweet text}}{\text{each field }\textit{sample}\text{ total tweet text}} \qquad 5\text{-}2$$

2. implementation

a) We implements this method in emotion analyze. First we set a set of key words which can represent a kind of emotion, then we use equation 5-2 to analyze each key word's reliability, if it is unreliable, we eliminate or replace it with another key word.

b) We implements this method in heat map. For each kind of exercise, we use its key word weight index multiple area's total tweet text, and use the result as total number of people in a area who actually doing exercise.

5.3.2 Personal suggestion based on vector space model (VSM)

Personal suggestion needs to find relationship between the user's information and our existing data and draw a conclusion. To fulfill this function, we use the VSM methods. The steps describe below:

a) With the help of GATE, both tweet texts(document) in our database and login user's information(query) are able to represented as vectors[13]

Each dimension represents a weight for term j in tweet text i, if the term showed up in the tweet text, then its value is none-zero.

b) Then we calculate the similarity of a document vector to a query vector using the cosine of the angle between them

particular, sim(d,q)=1 when d=q;

sim(d,q)=0 when d and q share no terms

From 5-4, we can define the similarity to the tweet text and user's information, then predict what kind of sports they will like based on other user's tweet text.

5.3.3 Sports correlation analyze based on linear regression

To analyze the relationship between two sports, we use linear regression methods to draw a line to show their correlation degree.

1. Linear regression math model description.

To implement linear regression analysis, we use generalized least squares methods (GLS)[14]. The roughly description shows below:

According to wiki [15], The GLS is applied when the variances of the observations are unequal, or when there is a certain degree of correlation between the observations. In our project, we assume the linear is:

$$\varphi(x) = a + bx \qquad 5\text{-}5$$

According to the GLS theory, the solution equation will be:

Within 5-6, m is the total number of point we use. From the solution equation, we can get a and b, then we put them back into 5-5 and get the linear equation.

To improve the accuracy, we introduce the Pearson product-moment correlation coefficient. According to the wiki [15], Pearson product-moment correlation coefficient is a measure of the linear correlation (dependence) between two variables X and Y. The mathematical equation is

$$r_{xy} = \frac{\sum_{i=1}^{m}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{m}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{m}(y_i - \overline{y})^2}} \qquad 5\text{-}7$$

The r_{xy} range from -1 to 1, we define $0.8 \le |r_{xy}| \le 1$ as highly correlated, $|r_{xy}| \le 0.3$ as no correlated.

2. Implementation

a) First we divide one day in to 24 hours and choose two sports that we want to compare, then we count how many related tweet text for each sport is showed in each hour. An example is showed as Figure 5-2.

b) Using equation 5-7 to analyze the data correlation degree, if they are highly correlated we use the generalized least squares methods to draw the linear equation in the coordinate and use the slope to judge if they are positive correlated or negative correlated.

	swimming	running
1h	500	1000
2h	250	600
3h	1500	2800
…	…	…

Figure 5-2. Example.

5.3.4 Sentiment analysis based on probability density function of a normal distribution

This method is derived from the project website of Visualizing Twitter Sentiment [16].

For sentiment analysis, the ANEW [17] dictionary provides measures of valence, arousal, and dominance for 1,034 English words. Each word is rated on a nine-point scale ranging from 1 to 9.

Ratings for a common word are combined into a mean rating and a standard deviation of the ratings for each dimension.

For example, for the word house, ANEW reports:

house,

$$v = (\mu:\ 7.26,\ \sigma:1.72),\ a = (\mu:\ 4.56,\ \sigma:\ 2.41),\ d = (\mu:\ 6.08,\ \sigma:\ 2.12),\ f_q = 591 \quad 5\text{-}8$$

This shows that house has a mean valence v of 7.26 and a standard deviation of 1.72, a mean arousal a of 4.56 and a standard deviation of 2.41, a mean dominance d of 6.08 and a standard deviation of 2.12, and a frequency f_q of 591 ratings.

However if multiple words documented in ANEW dictionary for instance:

Congrats to @HCP_Nevada on their health care headliner win!

ANEW's measure of the n = 2 words' means and standard deviations of valence and arousal are:

health,

$$v = (\mu:6.81,\ \sigma:1.88),\ a = (\mu:5.13,\ \sigma:2.35),\ f_q = 105 \quad 5\text{-}9$$

win,

$$v = (\mu:8.38, \sigma:0.92),\ a = (\mu:7.72, \sigma:2.16],\ f_q = 55 \quad 5\text{-}10$$

To combine the means for health and win, we assume that the individual ratings reported for each word form a normal distribution. Intuitively, if a word has a higher standard deviation, for example, a higher $\sigma_{v,i}$ for valence, the valence ratings for the word were spread across a wider range of values. If $\sigma_{v,i}$ were lower, ratings for the word clustered closer to the mean. Based on this, we use the probability density function[18] of a normal distribution to estimate the probability of the word's rating falling exactly at the mean.

Notice that if we'd simply used an arithmetic mean to compute the overall mean

valence M_v, we would have reported $M_v = (6.81 + 8.38)/2 = 7.56$. However, the standard deviation of valence for health ($\sigma_{v,1}$ = 1.88) is higher than the standard deviation for win ($\sigma_{v,2}$ = 0.92). Because of this, we weight win's mean valence $\mu_{v,2} = 8.38$ higher than health's mean valence $\mu_{v,1} = 6.81$. How we allocated the weight is explained below:

The normal distribution is parametrized in terms of the mean and the variance, denoted by μ and σ^2 respectively, giving the family of densities

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \qquad 5-11$$

Hence, the probability of the word health's rating falling exactly at the mean could be calculated by the formula above.

When $x = \mu_{v,1} = 6.81$, $\sigma_{v,1} = 1.88$, we derive $f(x) = \frac{1}{1.88*\sqrt{2\pi}} = 0.212$.

Now we allocate the word 'health' with weight $W_1 = 0.212$.

We could use the same steps to derive the weight of the word win $W_2 = 0.434$

Then we can calculate the overall mean:

$$M_v = (\mu_{v,1} * W_1 + \mu_{v,2} * W_2)/(W_1 + W_2) = 7.86 \qquad 5-12$$

Hence, the result is an overall mean M_v = 7.86 that falls closer to win's mean valence. A similar result can be seen for overall mean arousal M_a.

The probabilities are applied as weights when we sum the means. Using this formula, we compute an overall mean valence and arousal of: M_v = 7.86, M_a = 6.48

For other situations that multiple words documented in ANEW dictionary, we could use the similar way to combine them:

$$M_v = \frac{\mu_1 * W_1 + \mu_2 * W_2 + \cdots + \mu_n * W_n}{W_1 + W_2 + \cdots + W_n} \qquad 5-13$$

In Fig 5-3 it shows the result by using the above mathematical method. Mood value has been calculated for corresponding distinct exercise types.

type	mood_value
running	6.08273
cycling	5.76242
swimming	6.07984
basketball	5.96129
volleyball	6.52989
tennis	6.15888
football	5.56301

Fig 5-3 Exercise types and their mood value

6 Plan of Work

6.1 Project Management

Conference date and location:

Our team holds conferences twice a week on Tuesday and Thursday at the study room 1 in the Library of Science and Medicine.

6.1.1 Project basic structure work

Each project basic structure work is distributed to one or more team members. However, we do not suggest each team member take part into the whole basic structure work at the very beginning. The basic structure works are mostly fixed and low-layered, and they need little changed when develop the high-layered works – features development. This arrangement contributes to improve the basic structure setting up efficiency, and diminish the reproduce and conflict when several team members do the same work. Although each team member will only do some parts of the structure work, they need to know the whole picture. Thus each team member would report his work to every one at the conference. Besides, as pushing the work forward, the team members in different basic structure parts need to communicate with each other for the work integration. So actually, each team member realizes the whole picture of our work in the end.

Assignments	D. Yao	Z. Zheng	W. Zhang	Y. Wu	W. Fang	Y. Sun
Data Collecting	√				√	
Data Storing & Rearrangement			√	√	√	√
Website Display		√	√	√		√
IOS Display	√	√				
Management & Integration	√					

Table 6-1 Basic structure

In the Table 6-1, the target of data collecting is to get the public information, e.g. the users' non-private information and tweets from Twitter, and users' private information from a demography speculation API. The team members in the part also need to schedule the date to download data, and export and import the database

structure and data to other team members. Data storing and rearrangement is the design of the tables in the database. The team members in website and IOS display need to design the UI and display clear charts and maps to the viewer. The work for management is to collect creative ideas and breakdown works to each team member. The work for integration is to combine the work from each team member. The uniform rules of work are needed before the work distribution. Otherwise a lot of renaming works will appear later. The integration work also include writing communication files for JSON sending and receiving between the front-end work and the rear-end work.

6.1.2 Product ownership (Features)

Features	D. Yao	Z. Zheng	W. Zhang	Y. Wu	W. Fang	Y. Sun
Tweet heat in geographical distribution			√			
Tweet heat in variation tendency		√				
Tweet heat in exercising classification						√
Tweet heat in demography					√	
User ranking				√		
Exercising duration			√			
Exercising frequency	√					
Word frequency				√		
Correlation topics						√
Tweet sentiment		√				

Personal diagnosis	√					

Table 6-2 Feature distribution

In the Table 6-2, each team member is assigned with several feature works. That is, everyone in the team takes part into the data analyzing part – feature development. When a team member thinks out a new feature, he will report at the conference, then the work will be discussed and distributed.

6.1.3 Breakdown of responsibilities (Report)

Assignments	D. Yao	Z. Zheng	W. Zhang	Y. Wu	W. Fang	Y. Sun
Problem Statement & Reference	√					
Glossary of Terms				√		
Functional Requirements		√				
Non-Functional Requirements					√	
On-Screen Appearance Requirements						√
Stakeholders				√		
Actors and Goals				√		
Casual Description		√				
Use Case Diagram	√					
Traceability Matrix		√				

Fully-Dressed Description					√	
System Sequence Diagrams			√			
User Interface Specification						√
Domain Model		√				
System Operation Contracts					√	
Mathematical Model			√			
Plan of Work	√					
Layout	√					

Table 6-3 Report writing

In the Table 6-3, the writing works for each part of the full report 1 are distributed as shown. Although each part of the full report 1 is wrote by different team member, at the conference, each team member takes part into revising the report for the consistence.

6.2 Project Schedule

6.2.1 Before full report 1

Data Collecting (Server)	**14-09-12**	**14-10-16**
Investigate Twitter API	14-09-12	14-09-15
Investigate the database codes from the former groups	14-09-14	14-09-15
Re-establish the database from the former group 1 and refine the keywords	14-09-16	14-09-16
Discontinuously download the data by Twitter streaming API	14-09-17	14-09-22
Implement Twitter rest API	14-09-25	14-09-26
Investigate Facebook API, Google+ API and text	14-09-29	14-10-10

analytics APIs for demography information		
Figure out gender, age and type by text analytics API	14-10-11	14-10-16
Data Analyzing (Database)	**14-09-16**	**14-10-16**
Investigate the features in reports from former groups	14-09-16	14-09-22
Tweet heat > geographical distribution	14-09-23	14-10-09
Tweet heat > variation tendency	14-09-23	14-10-09
Tweet heat > exercising classification	14-09-23	14-10-09
Tweet heat > user ranking	14-09-23	14-10-09
Exercising duration	14-09-23	14-10-16
Personal Diagnosis	14-10-10	14-10-16
Data Displaying (IOS & Web)	**14-09-16**	**14-10-16**
Set up the communication between front-end and rear-end	14-10-07	14-10-09
Structure the IOS app	14-09-16	14-10-06
Implement the IOS app	14-10-10	14-10-16
Investigate map graphical API	14-09-16	14-10-09
Implement marker, heat and state maps	14-10-10	14-10-16
Implement all features from former groups	14-10-09	14-10-09
Display improves compared to former groups	14-10-16	14-10-16

Table 6-4 Past works

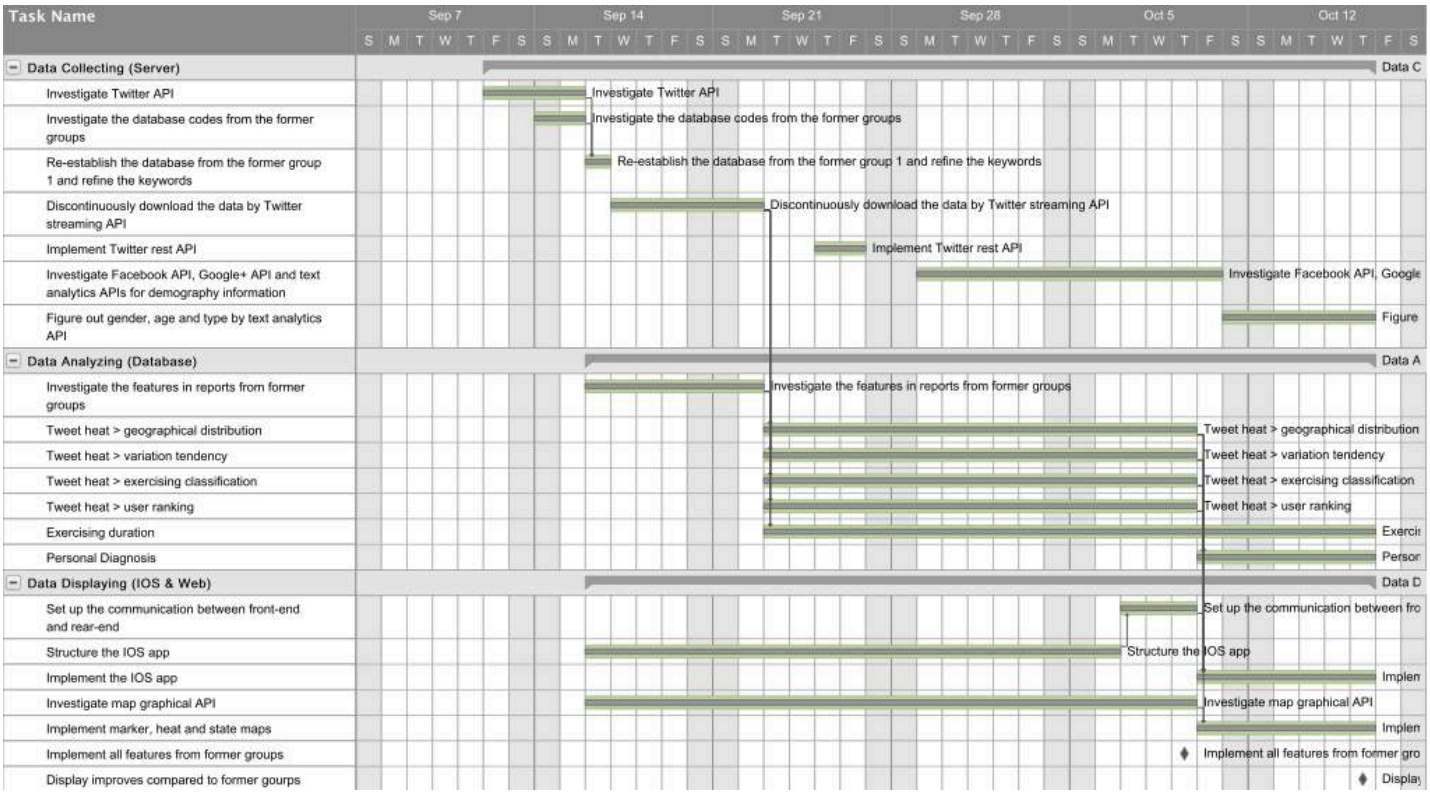


Figure 6-1 Past Works

6.2.2 After full report 1

Data Collecting (Server)	**14-10-13**	**14-12-05**
Download a full week data	14-10-13	14-10-20
Change access from private network to public network	14-10-21	14-10-30
Think about the new data	14-10-31	14-12-05
Data Analyzing (Database)	**14-10-16**	**14-12-05**
Personal diagnosis supplement	14-10-16	14-10-30
Word frequency	14-10-16	14-10-22
Topic correlation	14-10-23	14-10-30
Exercising frequency	14-10-21	14-10-30
Tweet mood	14-10-16	14-10-30
Demography	14-10-16	14-10-30
Share the personal diagnosis	14-11-01	14-11-07
Refine the implemented features	14-10-29	14-12-05
Think about new features	14-10-29	14-12-05
Data Displaying (IOS & Web)	**14-10-16**	**14-12-05**
Refine the IOS app	14-10-16	14-10-31
Structure the website	14-10-16	14-10-31
Implement the website	14-11-01	14-11-20
Refine the website	14-11-21	14-12-05
Implement lots of new features	14-10-30	14-10-30
First demo	14-10-31	14-10-31
Second demo	14-12-05	14-12-05

Table 6-5 Future plans

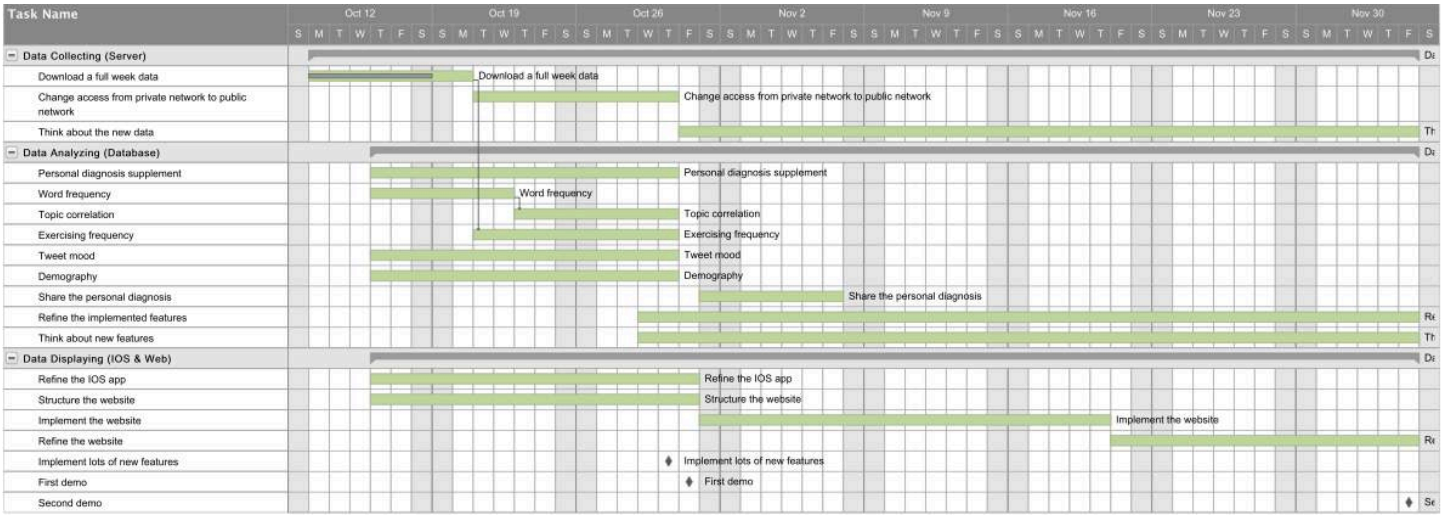


Figure 6-2 Future plans

References

[1] *HCA*: http://en.wikipedia.org/wiki/Health_care_analytics
[2] *HMA 2014 project description*: http://www.tru-it.rutgers.edu/takmac/
[3] S. Fox, M. Duggan. *Health online 2013*. Jan. 15, 2013.
http://www.pewinternet.org/2013/01/15/health-online-2013/
[4] S. Bennett. *Facebook, Twitter, Instagram, Pinterest, Vine, Snapchat – Social Media Stats 2014.* June 9, 2014.
http://www.mediabistro.com/alltwitter/social-media-statistics-2014_b57746
[5] *Phirehouse*: https://github.com/fennb/phirehose
[6] S. Kumar, F. Morstatter, H. Liu. *Twitter Data Analytics*. Aug. 19, 2013.
[7] *JSON*: http://en.wikipedia.org/wiki/JSON
[8] *Twitter API*: https://dev.twitter.com/overview/documentation
[9] *Rest API*: https://github.com/abraham/twitteroauth
[10] *Demographic API*: http://textalytics.com/core/userdemographics-info
[11] PNchart: https://github.com/kevinzhow/PNChart
[12] *Google API*: https://developers.google.com/apis-explorer/#p/
[13] *The VSM model*:
http://www.csee.umbc.edu/~ian/irF02/lectures/07Models-VSM.pdf
[14] *The least square method and accuracy analysis*:
http://wenku.baidu.com/link?url=gwzOkBt_AQSSeWbnytidM9qEQT007jsxqC9Uqpt7B5qSUPBZicnFyQGs2LRFwPrr8zvaR0PmA9nR0uOVKUvpj8s60PDma8mYILzbK617wyq
[15] *Pearson product-moment correlation coefficient*:
http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient
[16] *Healey & Ramaswamy. Visualizing Twitter Sentiment*:
http://www.csc.ncsu.edu/faculty/healey/tweet_viz/
[17] Margaret M. Bradley and Peter J. Lang. *Affective Norms for English Words (ANEW).*
[18] *Probability density function*:
http://en.wikipedia.org/wiki/Probability_density_function