

CERTIFICATION OF APPROVAL

**Web Based Disease Data
Visualization and Interpretation System**

by

Chow Cheah Hui

A project dissertation submitted to the
Information & Communication Technology Programme
Universiti Teknologi PETRONAS
in partial fulfilment of the requirements for the
Bachelor of Technology (Hons)
(Information & Communication Technology)

Approved by,



(Ms. Goh Kim Nee)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
January 2011

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.



CHOW CHEAH HUI

Acknowledgement

First of all, I want to express my gratitude to my supervisor, Ms. Goh Kim Nee who had provide me the initial idea of this project and guide me along the way. I also appreciate Mr. Deva, head of Data Surveillance Unit of Jabatan Kesihatan Daerah Kinta for providing me the disease data and also help to test my system. In addition, I would like to express my immense gratitude to Universiti Teknologi PETRONAS and to my lecturers.

This project would not have been possible if not for the generous assistance of many individuals who shared their knowledge and expertise. To all of you, I extend my utmost gratitude and deep appreciation.

TABLE OF CONTENTS

ABSTRACT	vi
CHAPTER 1:INTRODUCTION	1
1.1 Background of Study	1
1.2 Problem Statement	2
1.3 Significance of the Project	3
1.4 Objectives and Scope of Studies	4
CHAPTER 2:LITERATURE REVIEW	5
2.1 Impact of Environment Factors on Vector Borne Diseases	5
2.2 Using Mathematical Models to Predict the Trend of Diseases	5
2.3 Data Visualization	6
2.4 Projects that use Data Visualization	8
CHAPTER 3:METHODOLOGY	13
3.1 Research Methodology	13
3.2 Data Collections Methods.	15
3.3 Tools used	19
CHAPTER 4: RESULTS AND DISCUSSION	20
4.1 Use Case Diagram	20
4.2 Activity Diagram	21
4.3 Database Design	22
4.4 Display Map	26
4.5 Discussion	34

CHAPTER 5: CONCLUSION	35
REFERENCES	36
APPENDICES	39
1. Key Milestones	40
2. Project Gantt chart	41
3. Letters to request for data	43

LIST OF ILLUSTRATIONS

Figure 1	CrazyEgg	9
Figure 2	Map visualization of disease cases	10
Figure 3	The disease historical data	11
Figure 4	Prototyping Methodology	13
Figure 5	Google Insights	17
Figure 6	Dengue Disease Weekly Report	18
Figure 7	Use Case Diagram	20
Figure 8	Activity Diagram	21
Figure 9	Markers	22
Figure 10	Raw Tuberculosis Disease Data	23
Figure 11	Append the state and country	24
Figure 12	Before run geocoding	25
Figure 13	After run Geocoding	26
Figure 14	Tuberculosis for year 2007	27
Figure 15	Tuberculosis for year 2008	28

Figure 16	Tuberculosis for year 2009	29
Figure 17	Tuberculosis for year 2010	30
Figure 18	Choose start and end date	31
Figure 19	Map that shows only diseases within user queried date	32
Figure 20	Choose start and end date for Interpretation	33
Figure 21	Interpretation	33

ABBREVIATIONS AND NOMENCLATURES

CSV	Comma delimited Values
GIS	Geographic Information System
GPS	Global Positioning System
2D	Two Dimensions
3D	Three Dimensions
J2EE	Java 2Platform Enterprise Edition
UTP	Universiti Teknologi PETRONAS
HFMD	Hand, foot and mouth disease
HIV	Human Immunodeficiency Virus
AIDS	Acquired immune deficiency syndrome
TB	Tuberculosis

ABSTRACT

Malaysians suffer from both communicable and non-communicable diseases. However, an easy to use tool is not available for the doctors, officers from Ministry of Health and also the public to analyze the disease. The data is not centralized and even if the user can collect sufficient data to analyze the data, many people do not have a clue about the overwhelming data. This project will develop a website that will visualize the disease data on a map stating the location that the diseases happened. As everyone might interpret the result differently, a paragraph of interpretation will be generated along with the visualization to give an impartial description about the data. This research consists of all the dengue and Tuberculosis cases in Daerah Kinta. Tuberculosis disease is chosen because it is more common in rural places like some parts of Daerah Kinta. On the other hand, Dengue is chosen because it is a very common vector borne disease in Malaysia. The methodology used is prototyping as it allows more users' feedback in the early stage of the system. Mock ups will be generated to allow users to interact with it. The map is shown by using Google maps API. The author will use Google Geocoding API to convert the addresses where diseases happened into longitude and latitude to mark it on the map. The interpretation is generated by comparing the number of disease case in one period of time in an area with the medical standard provided by Jabatan Kesihatan Daerah Kinta.

CHAPTER 1

INTRODUCTION

1.1 Background of Study

In 2007, the top five diseases in Malaysia were dengue fever, tuberculosis, food poisoning, hand food and mouth disease (HFMD) and HIV/AIDS. The incidence rates were 80.6 per 100 000 population for dengue fever, 61.9 per 100 000 for tuberculosis, 52.6 per 100 000 for food poisoning, 46.1 per 100 000 for HFMD, and 16.0 per 100 000 for HIV/AIDS [1]. Dengue fever which is a vector borne disease topped the rank. All of the top five diseases are communicable diseases which mean it can be transmitted to other human through direct or indirect contact.

Other than the usual diseases, pandemic influenza is also threatening the health of not only Malaysians but also many people around the world. Pandemic influenza arrives with little warning and can cause huge damage in a short time. The risk of influenza had also increased as the mobility of the population becomes greater [2].

Early prevention of the diseases will save the money spent on combating the diseases later. According to Guy Richard, mathematical models can predict the way the communicable diseases will spread and possibly prevent them from spreading [3]. Google Flu Trend is a project by Google to estimate the flu activity. It is said that the number of people query for flu related topics using Google search and the number of people experiencing flu symptoms are closely related. When compared the popularity of flu related queries with the official data that were released two weeks later than the queries statistic, the numbers actually matched although there is some small variations [4].

There is also an online application by Google called Google Insight that allows the public to search for the popularity of any query topics from the year of 2004 to now from any part of the world. Instead of displaying the real search volume of query keywords, Google Insights only display the normalized data in the graph that shows the popularity of a keyword [5]. Since every area has different population, if Google don't normalize the data, the area with relatively higher population will always show as the top interest in any keywords. As an example, let say

there are 6000 Internet users in Penang state and 3000 of them query for the keyword 'Panasonic'. In Pahang, a much bigger state might have 20000 of Internet user and 5000 of them queried for the keyword 'Panasonic'. We can see that in Penang the percentage of people interested in Panasonic are higher than the percentage of people interested in 'Panasonic' in Pahang. However, if Google does not normalize the data, it would appear as Pahang people are generally more interested in 'Panasonic' and might lead to wrong decision made by the company using Google Insights to survey the market.

The spread of some diseases are affected by the climate change. As an example, the cases of some vector borne disease such as Dengue and Malaria increase at certain climate condition that is suitable for the breed of the mosquitoes. The cases of some respiratory disease such as Asthma are affected by the humidity of the air too [6]. Since weather data can affect the occurrence of certain disease, it is important to include these data in the visualization project.

1.2 Problem Statement

The outbreaks of a new disease always remind the public about the Spanish flu pandemic of 1918-19 which is estimated to have killed 40 million people worldwide. This perception have caused panic and thus weakened the world economy. The areas that are adversely affected by this situation are air travel, tourism, and meat imports and exports. This is due to the fact that people try to avoid catching the disease by postponing their travel plan. The countries also grow to be more careful when import foods from foreign country. However, the situation still continues although the World Health Organization already relaxed the travel restrictions that are recommended earlier and meat is proven to be not a medium where virus can spread. The circumstance above shows that the panic caused by the virus seems to be more destructive than the virus itself. In order to overcome this problem, it is crucial for us to analyze the spread of a virus so that action proper to the risk level of the virus can be taken [3]. However, in order to analyze the spread of the disease, sufficient information would be needed. World Health Organisation has some information on communicable diseases but it is on annual basis. The hospital has more up to date information but the information is not publicly available. It requires formal communication such as letter from a university with acceptable reason(s) to

require the data. As a result, it is very troublesome for the public to obtain the sufficient information and made a good analysis.

1.3 Significance of the Project

According to the Health Facts 2007 published by the Ministry of Health Malaysia, one of the strategies that will guide the health sector development is: Enhancing research and development to support evidence-based decision-making. This project aims to provide a mean for doctors, officers from Ministry of Health Malaysia and also the public to understand and relate the data related to diseases. The ability to understand such thing will assist them in decision making process.

Malaysia Health Profile published by World Health Organization shows that 28% of Malaysian passed away due to communicable diseases during 2008 [1]. This shows that if we are able to educate the public to prevent the communicable diseases to spread by taking some precaution steps, the number of death and also the money spent on curing the diseases can be reduced tremendously. The existence of system which visualizes the disease data on a map and generate an interpretation will certainly make it easier to learn about the disease and thus lower the possibilities of catching the communicable disease by avoiding travel to the high risk area. Since the users can choose the area and the duration of the disease spreading to view, this project can also assist them in analyzing the data over time and finding out the impact of weather on diseases.

Questionnaires has been distributed to 65 individuals who are currently studying or working in Malaysia, out of 24 responses, more than half of them are not aware of the number of cases of communicable diseases in their area. 87% of the respondents agree that it is easier for them to analyze the disease data if there is visual aid with a short paragraph describing the visual aids. 79% of them find it helpful if the disease spreading tendency is represented in a map. Moreover, 79% of the respondents said that if there is a website providing a map of Malaysia with the disease spreading tendency, they would you accesses it. (Refer to Appendix for full result.) Based on the result of the questionnaire, it is clear that the respondents think that this project can assist them to analyze the disease data.

From the top five diseases in Malaysia in the year of 2007, the author realize that most of the disease can be prevented if the public has awareness and take action to prevent the diseases itself. For example, if everyone spend a little time to check their surrounding areas and throw away empty cans, bury anything that can hold water etc, the mosquitoes will not have place to breed in that area. To be safe, the residents can also put on insect repellent cream to prevent getting bitten by mosquitoes.

1.4 Objectives and Scope of Studies

1.4.1 Objective

According to Palaniappan, S. and Awang, R., many useful information can be retrieved if the data contained in the health care industry can be processed and analyzed [7]. The author aims to assist the analysis of disease data and weather data in Perak state by plotting the information on the map where the disease happens. A short paragraph of the result interpretation will be generated to provide an impartial description of the data presented on the map.

1.4.2 Scope of Study

This project will produce a website that visualizes the disease data and weather data on a map and accompanied by a paragraph of description interpreting the visualization result. The data will be manipulated before it is being fed to the system if necessary. The author chooses to develop software to visualize the data as it is easier to interpret an image rather than text [8]. Visualization reduces the effort to understand data and data mining results, and allows comparisons and testing to be done on the result. It will also made the data mining algorithms more understandable to the users and thus enable them to be involve in decision analysis [9]. The scope of this project will be limited to certain areas in Perak only because of the time constraint of this project.

As different people might interpret the visualization on the map differently, a paragraph of interpretation will be generated with the visualization. This paragraph will be generated based on the benchmark that is established from variation of the data over two years so it will be more impartial.

CHAPTER 2

LITERATURE REVIEW

2.1 Impact of the Environmental Factors on Vector Borne Diseases

Malaysia has high number of cases of dengue fever and malaria, both are vector borne diseases. Webhwa ZENG et al. assessed impact the environmental factors have on vector borne diseases and conclude that the relationship is valid. However, it is mentioned that the remote sensing and GIS method used in identifying and monitoring the environment factors will not be a reliable epidemiological technique if the understanding of the relationship between the factors are not sound[10]. Previous UTP students have also done their research to find out how GIS factors in Batu Gajah area affecting the number of Dengue cases in the year of 2006 and 2007. They use the software MapInfo Map to create and edit a map. MapInfo also allow them to embed the map in their word processor and edit it using MapInfo functions. They collect the GIS information and put it in their map. After that, they compare it with the suspected dengue cases data that they obtained from the Batu Gajah Hospital. Hypothesis such as the mosquitoes bite the victim when they are walking on the road, when they stay nearby to the drain and when they stay in their house with a toilet bowl (mosquitoes breeding place). These hypotheses are proven to be true since all the dengue cases that happened at least fulfilled one of the assumptions. [11][12]

Studies found that temperature will affect the population of mosquitoes in thus the incidence rate of dengue and other vector borne disease. The change of climate due to global warming may increase the population of mosquitoes and results in more dengue cases. The area that used to be dengue safe might be affected by this factor [13]. It is very important to monitor the factors that effect the spreading of the diseases so that it can be combat.

2.2 Using Mathematical Models to Predict the Trend of Diseases

Guy Richard mentioned in his article 'Don't Panic!' that mathematical models can predict the way diseases and viruses will spread. According to Guy, the more data we can gather the better it is because more data permit us to use simpler models. However, in this imperfect world, the data is usually incomplete. As a result, lots of models are used together by epidemiologists to

reach to an agreement. An important quantity that is used by all the models is basic reproduction number of a disease- R_0 . This is the number of other people a person who is infected with a disease will infect. If it is less than one, the infection will die out, if it is equal to one, the disease will remain in the population at a same rate, if it is more than one, it is epidemic. However, since the R_0 of the diseases is derived from the previous outbreak of the disease, it might not be accurate as the disease will vary from time to time. There are two types of model which are deterministic and stochastic. Deterministic model assume everyone susceptible to the disease at the same rate and the disease will spread in a fix rate according to the level of contact between humans. Stochastic model utilize random variables and probabilistic factors to determine the way diseases will transmit as it assume everyone is susceptible in different level [3]. Models that based on historical data can be apply to detect the influenza activity as influenza is recurring disease that already have cases of infections[14].

Due to the time constraints of this project, there would be no mathematical models used to predict the trend of the communicable diseases in this project. However, the disease trend prediction will be a great future enhancement for this project. Moreover, in order to do this project, the author needs to collect the disease data and weather data. This data can be fed to the mathematical model to generate the prediction result to further enhance the website.

2.3 Data Visualization

2.3.1 What is Data Visualization?

The word 'visualize' means to make a mental image of; and also to make visible [15]. Visualization is to present the information in graphical format to enhance the qualitative understanding of the information from the viewers [16].

According to Friedman (2008),

"The main goal of data visualization is to communicate information clearly and effectively through graphical means. It doesn't mean that data visualization needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand,

providing insights into a rather sparse and complex data set by communicating its key-aspects in a more intuitive way. Yet designers often fail to achieve a balance between design and function, creating gorgeous data visualizations which fail to serve their main purpose — to communicate information [17]".

The main goal of data visualization is to communicate the data clearly, and in order to know whether the data has been communicated clearly, we need to ask the opinion from the users.

2.3.2 Visualization Techniques

Two types of goals for visualization techniques are geometric and symbolic. When deciding the visualization technique, we also need to determine whether the stimulus will be 2D or 3D, and also decide to have either dynamic or static stimulus.

Examples of data representation using geometric techniques are scatter plot, lines, surfaces or volume. Most of the time this types of representation involves some columns of the data being mapped to the axes of the graph that are being displayed.

On the other hand, symbolic representations are usually used to display the relationship between some elements of the data. Symbolic representations use pixels, icons, array or graphs (network relation etc) to display the data. Both geometric and symbolic representation can be presented as 2D or 3D, hence it is more important to decide the goals of the representation first.

Lastly, it is also equally important to decide whether the representation should be dynamic or static [18]. Dynamic representation will be more interaction but it could be more complex.

2.3.3 Requirements for a good visualization

Below are the general requirements for good data visualization [16] [19]:

- **Aesthetics** – The visualization must look good to the eyes so that the viewer will feel good looking at it
- **Effective** – The visualization must be able to communicate the data so that the viewer can understand it easily.
- **Adaptable** – The visualization should be able to be adjusted to serve different needs.

2.4 Projects that use Data Visualization

There are several interesting projects that use data visualization to communicate ideas.

2.4.1 CrazyEgg [20]

Crazy Egg is an application using visualization techniques to show the how the popularity of sections to blog visitors in a heat maps. The section that has most click from the visitors will be shown as the warmest color (red). Figure 1 below shows an example of screen viewed using CrazyEgg.

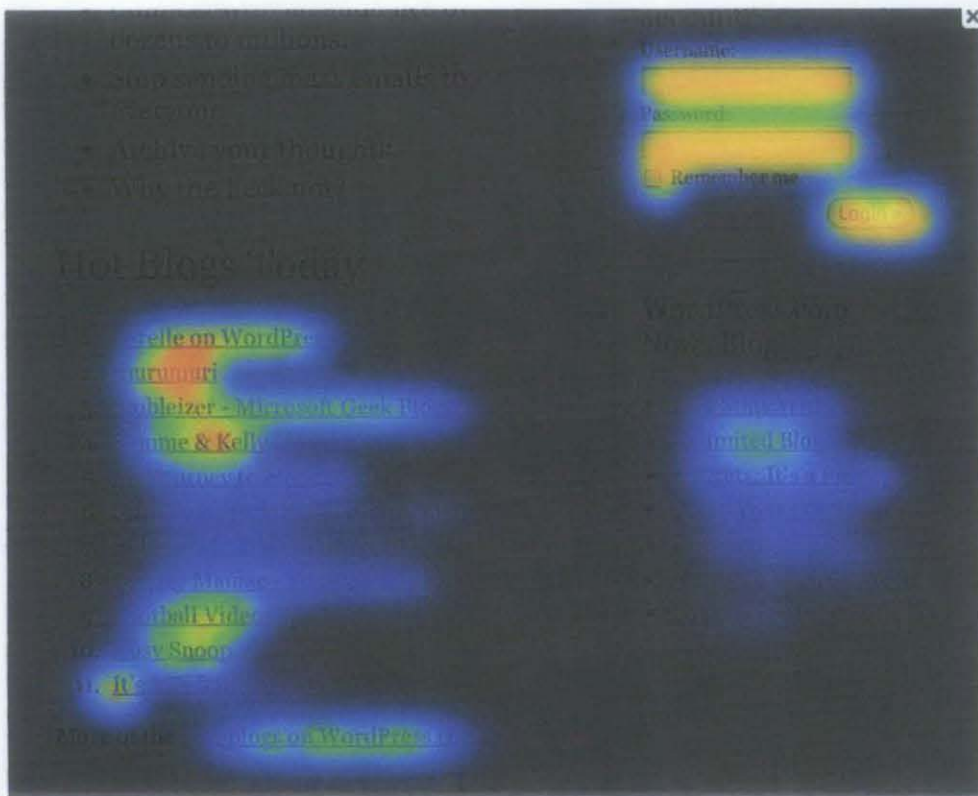


Figure 1: CrazyEgg

2.4.2 Web GIS based Information Visualization for Infectious Disease Prevention [21]

This project visualizes the number of disease case in different state in China using color. Different color represents different number of cases (Figure 2). If the viewer wants to know more about the trend of the disease over time, the viewer can click on the map. A visualization of the trend of the disease over a period of time will be shown (Figure 3).

Figure 2 below shows the visualization of the disease developing tendency in a map.

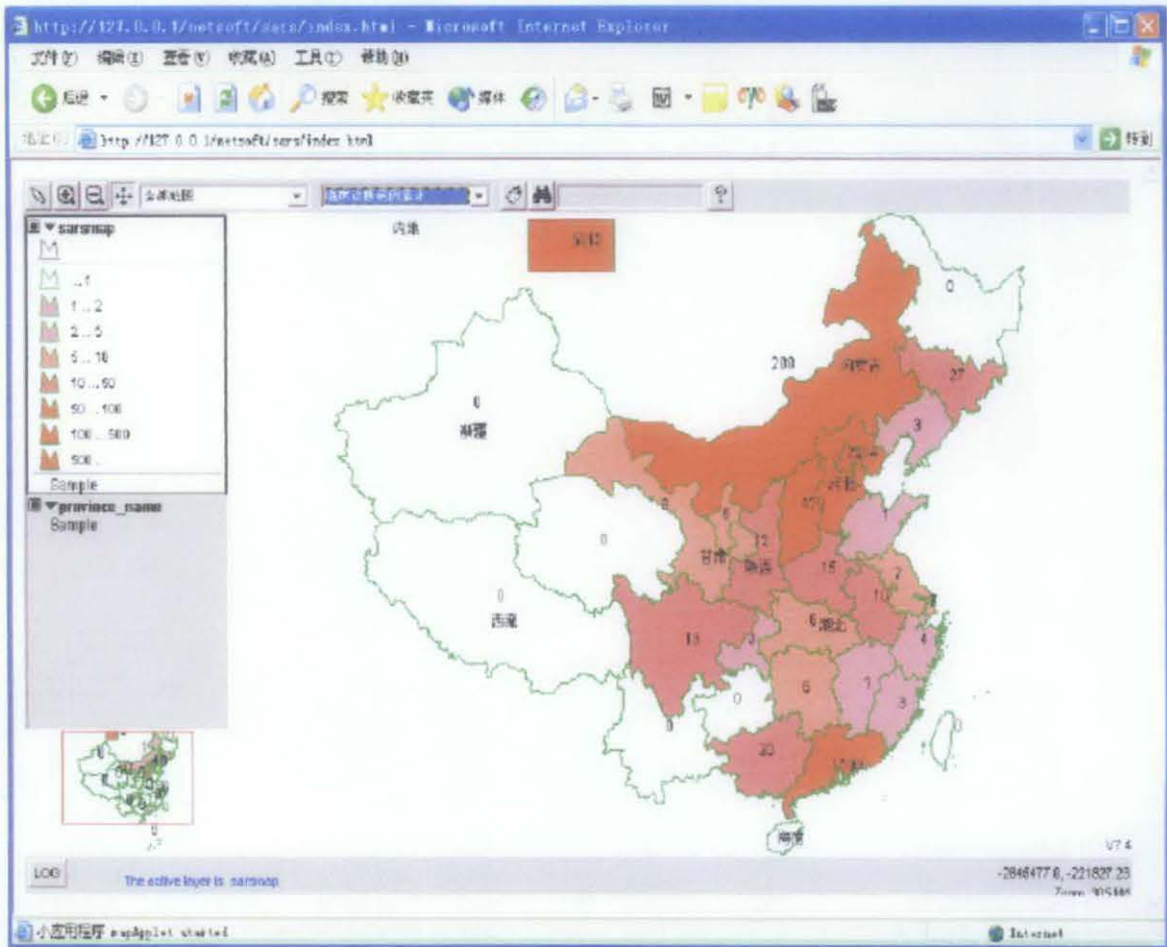


Figure 2: Map visualization of disease cases

Figure 3 below shows the visualization of the number of person infected with the disease, passed away from the disease as well as recovered from the disease over a period of time.

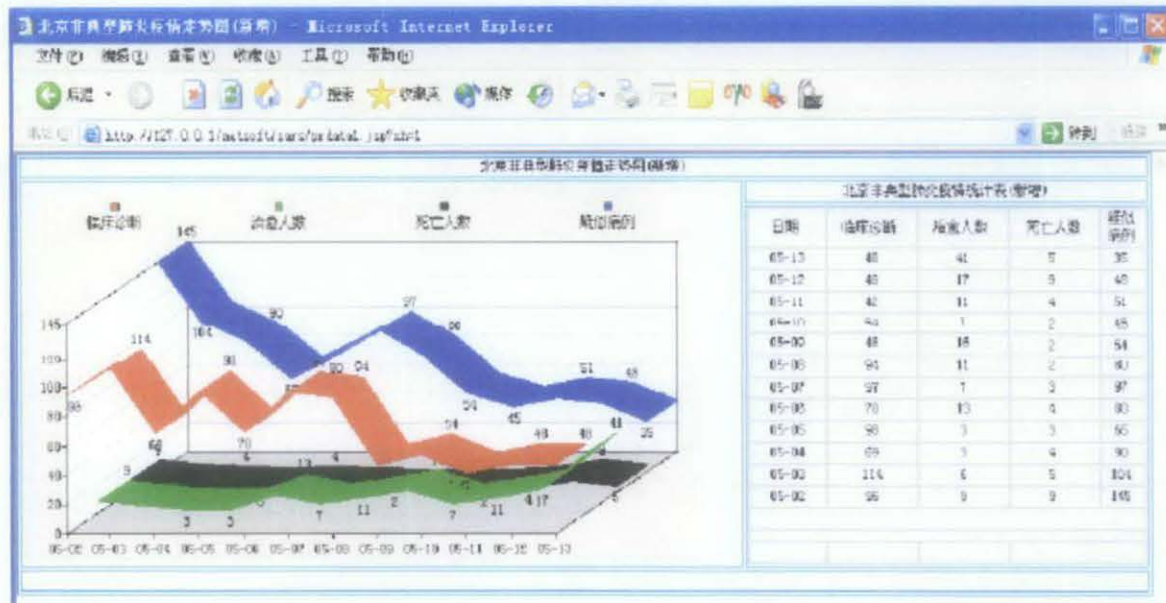


Figure 3: The disease historical data.

This is a web based project that provides the visualization in the Internet. J2EE based architecture is used in this project for the distributed system infrastructure. This project is built using a combination of GIS Technology, visualization technology, GPS technology and Web technologies. First, the disease data, medical information, graphics data, image data will be integrated to form a system database. The early warning disease surveillance system can be established from this system database. Based on the J2EE framework, there are server end and client end for this platform. The client end will be used to test the functions of the user interface.

GIS and Public Infectious Information Visualization

When the partial data of GIS and the infectious disease information are combined the infectious disease distribution information can be communicate through digital map.

Information such as number of infected person and the source of the virus can be presented in the map.

Information Visualization with Distributed Maps

The fact that this project combines disease data with medical information and GIS data allows the project to be used to track the infectious disease spreading tendency. Other than that, it can also be used to find out the best route to use for the personnel to travel in the infectious area.

CHAPTER 3

METHODOLOGY

3.1 Research Methodology

The research methodology that will be used for this project is Prototyping. The author chooses to use prototyping methodology for this project because it allows the author to constantly made changes to the existing program after the user review. The author can go through the phase of analysis, design prototypes, and implementation many times. Prototyping also allow user's feedback at the early stage of the project, this will make sure the author align with the objectives of the project from the beginning.

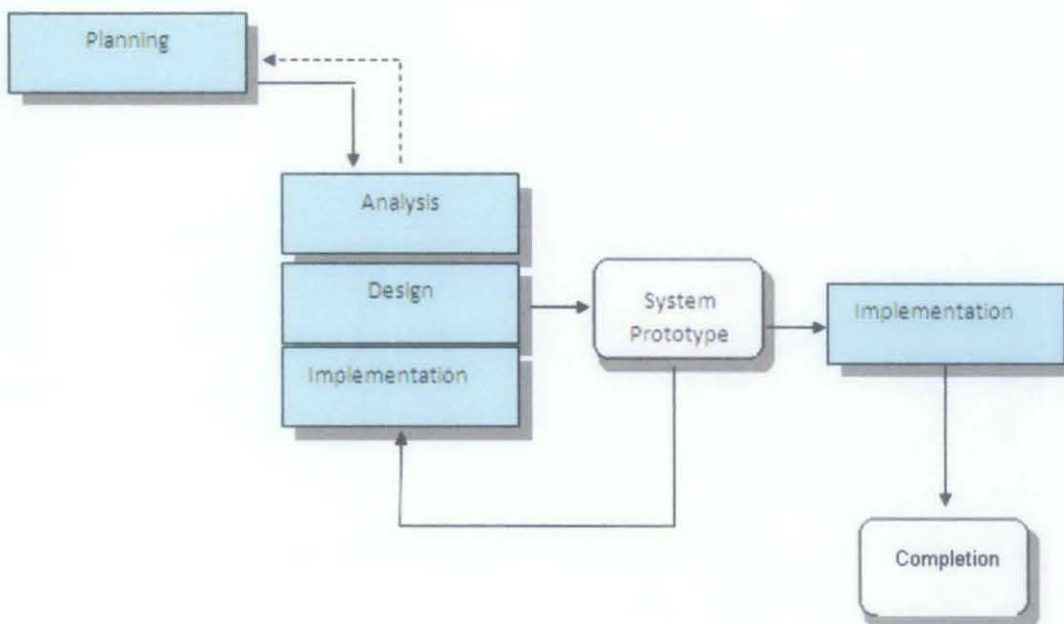


Figure 4: Prototyping Methodology

3.1.1 Activities in Each Stage

Planning

In planning stage, the author will conduct research about the project topic. Literature review will be written to discuss the technology used in other project similar to this project. Books related to this topic such as visualization and data mining will be referred to as well. Other than that, the author will try to collect data that are necessary for the project such as communicable diseases data for a reasonable period of time. The author will try to find reliable source for these data and will based the future stage of the project according to the format and availability of the data collected.

Analysis

During this stage, the author will identify the requirements of the users of the system. The user consists of doctors and also public who are interested in the trend of disease. In analysis stage, the author will also create a use case diagram to communicate what user can do with the developed program. An activity diagram will also be done to communicate the flow of this system.

Design Prototype

In the design prototype phase, the author will perform the Data Management layer design. The author will use relational database to store the data. The Data Management layer design will design the tables with the columns in the tables. The way where the tables will connect to each other will also be included in this design.

Implementation

The author will start coding according to the design to construct the prototype. First, the database needed to be build based on the design and the data collected. After that, the author will write a program to plot the disease according to the address in the database to the area in

the Google map that will be included in the website. The prototype will be tested using unit test after a single function has been done. Integration test will be conducted after combining all the functions together as a system. This is to make sure that all the functions can produce the same result when they combined as when they are being run individually.

System Prototype

The prototype will be presented to the users including the public and doctors to get their feedback. If the users approve the prototype, the author will complete the system components and install prototype. However, if the users do not approve it, the author will need to note down the feedback and redo the analysis phase by including the new requirements. The project will move from analysis phase to other phase again, this time with the changes based on users' feedback. The process will continue until the users approve it.

Implementation

If the users approve the prototype, the author will complete the components based on the prototype that is approved by the users. Testing such as system test, user acceptance test will be conducted at this stage. The result of the visualization will be analyzed, if there is some interesting relationship found between the variables from the visualization, doctors or some experts will be interviewed to discuss about the result.

Completion

The system will be completed and improvement suggestions will be taken for future enhancement.

3.2 Data Collection Methods

As this project depends heavily on data, the data collection methods are very important. Without data, the author cannot manipulate the data using the software that she intends to develop and cannot analyze the result. Several data collection methods used in this project will be discussed below:

3.2.1 Kinta Health Centre

The author managed to get the Tuberculosis and Dengue data from the Jabatan Kesihatan Daerah Kinta (Kinta Health Centre). Tuberculosis is chosen by the author because the Mr. Deva, a health officer from Data Surveillance Department of Jabatan Kesihatan Daerah Kinta suggested that it would be useful to mark this disease on the map since this disease is contagious and it is quite common some rural parts of Daerah Kinta. Daerah Kinta covers Tanjung Rambutan, Tronoh, Chemor, Gunung Rapat, Gopeng, Kampar, Malim Mawar, Menglembu, Pasir Pingi, Manjoi, Ulu Kinta and Ipoh. As a result, the Dengue and Tuberculosis data will be the cases that happened on those areas only.

3.2.2 Internet

3.2.2.1 Google Insights [5]

Google Insights is a web application developed by Google that allows the public to retrieve the popularity of any Google Search keywords in any part of the world where there are people using Google Search Engine. The duration of the data can be as early as 2004 until recently. The data can also be downloaded in .csv file which can be open using Microsoft Excel. This will make it easier to utilize the data from which the author can insert the data to database and manipulate it. The graph of Google Insights are actually showing the normalize data using this equation: $\text{normalized amount} = \text{amount of people query the keywords} / \text{amount of people using Google search over time}$. The below is an example of using Google Insights to find the popularity of the keyword 'Dengue' in Malaysia for the year of 2009 and 2010.

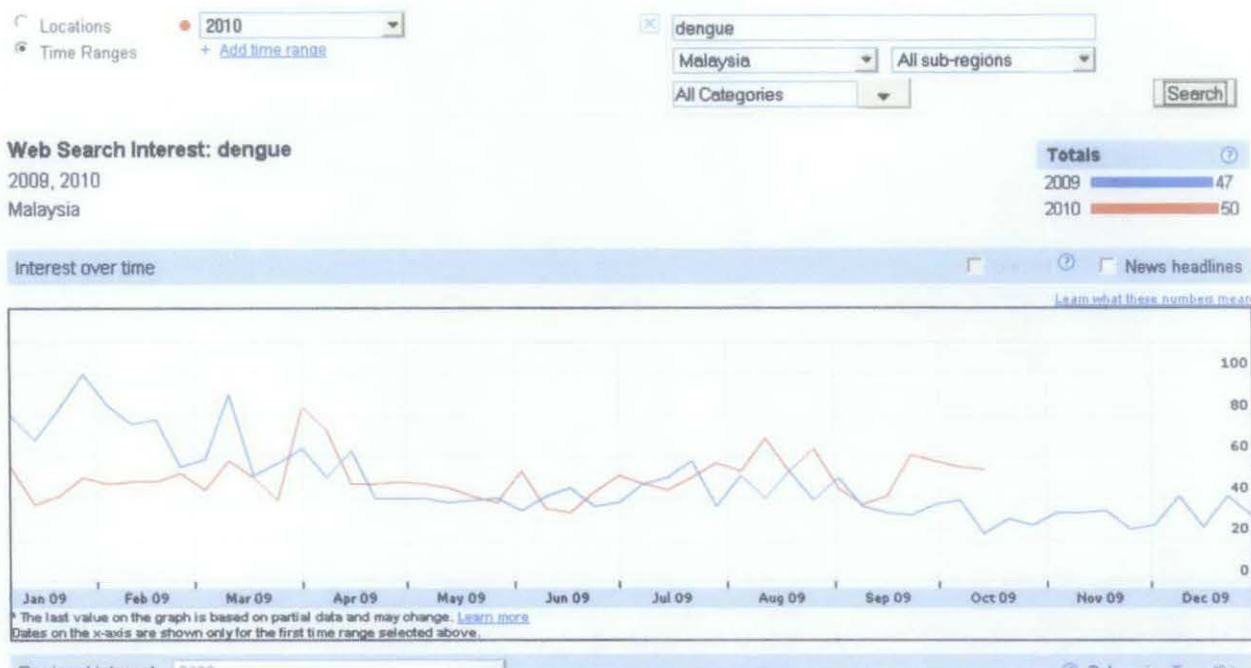


Figure 5: Google Insights

We can observe that the amount of Google search for the keyword 'Dengue' in the beginning of the year of 2009 is more than the amount in the beginning of 2010. Let's see the real statistic of cases of Dengue in Malaysia during these two years. This graph is taken from the Ministry of Health, Malaysia.

DENGUE DISEASE WEEKLY REPORT

(Up to Epidemic week 19/2010: 9 – 15 May 2010)

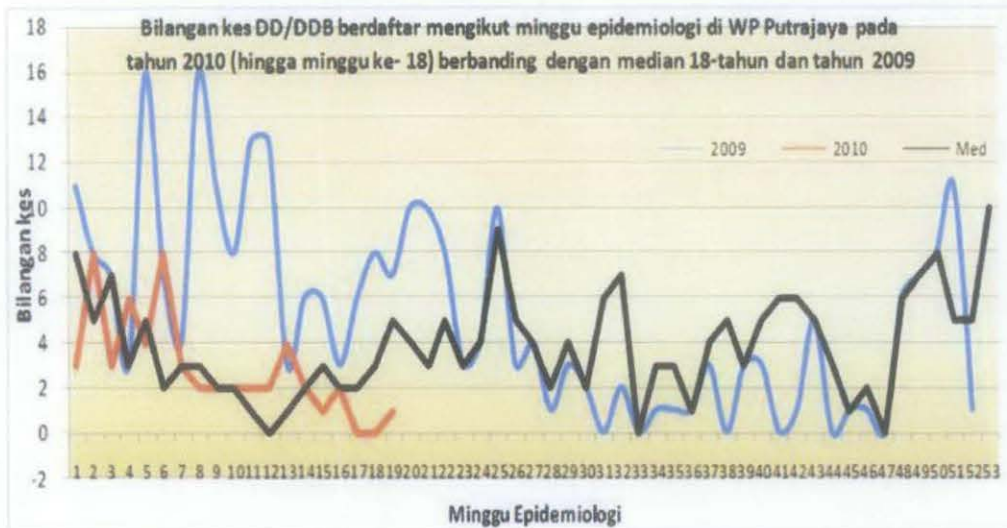


Figure 6: Dengue Disease Weekly Report [22]

(The x-axis shows the epidemiology week while the y-axis shows the number of cases.)

The amount of search in the beginning of the year of 2009 and 2010 for the Google Insights and the real statistic actually match. However, the trend starts to vary in the mid 2009 and mid 2010. The intention of putting this example in here is to show the reader how the comparison can be made. Since it is more tedious to make the comparison like this, the author is motivated to develop a system which can combine all the data and visualize it in a standard scale.

However, there is one disadvantage of Google Insights data. The graph cannot be shown when the Search volume is too little. Since the author plan to do the project in Perak, the author had tried to use Google Insights to search for some diseases in Perak state. The result is all of the keywords have too little search volume to show the graph. When the volume is too little, the Excel file downloaded will shows zero as well.

As a result, Google Insights will only be used as input when the search volume is high enough.

3.2.2.2 Malaysia Geographic Information System

Based on the literature review in section 2.4.2, the GIS data when include in the visualization can make the visualization much more useful. For this reason, the author will get the map of the research area from Google Map so that the disease data can be plot on the map based on the area.

3.3 Tools Used

3.3.1 Google Maps API Geocoder

Geocoding is the process of converting geographic related data such as street names and zip codes into a set of latitude/longitude coordinates so that it can be plotted on a map. Google Maps API Geocoder is a free geocoding service provided by Google [23]. It is very useful in such a way that the author can send the address where the disease happen to the Google Maps API Geocoder and get the longitude and latitude of the address. Thus, the location can be marked on the Google Map embedded in the proposed system's website. The author will use and PHP and JavaScript to develop a program that utilize the Google Maps API Geocoder to plot the map for disease cases and also to color the map for the weather visualization.

3.3.2 Database Tool and PHP server

Since this project requires visualization of data, it requires a database tool to store and appropriately manipulate the data before it is visualize. Possible way of manipulation might be cleansing the data to eliminate dirty or incomplete data if there is any. The author will use phpMyAdmin as a database tool for websites because it is open source and the author has experience using it. The user also downloads Apache so that PHP file can be run. To make it simple, the user just install XAMPP that contains everything the author need.

CHAPTER 4

RESULT AND DISCUSSION

4.1 Use Case Diagram

Below is the use case diagram of the system. The users choose the area and the duration of the data that they wish to view. They also need to choose the disease type that they wish to view. Then they will be able to view the map visualization and also a paragraph of interpretation of the data.

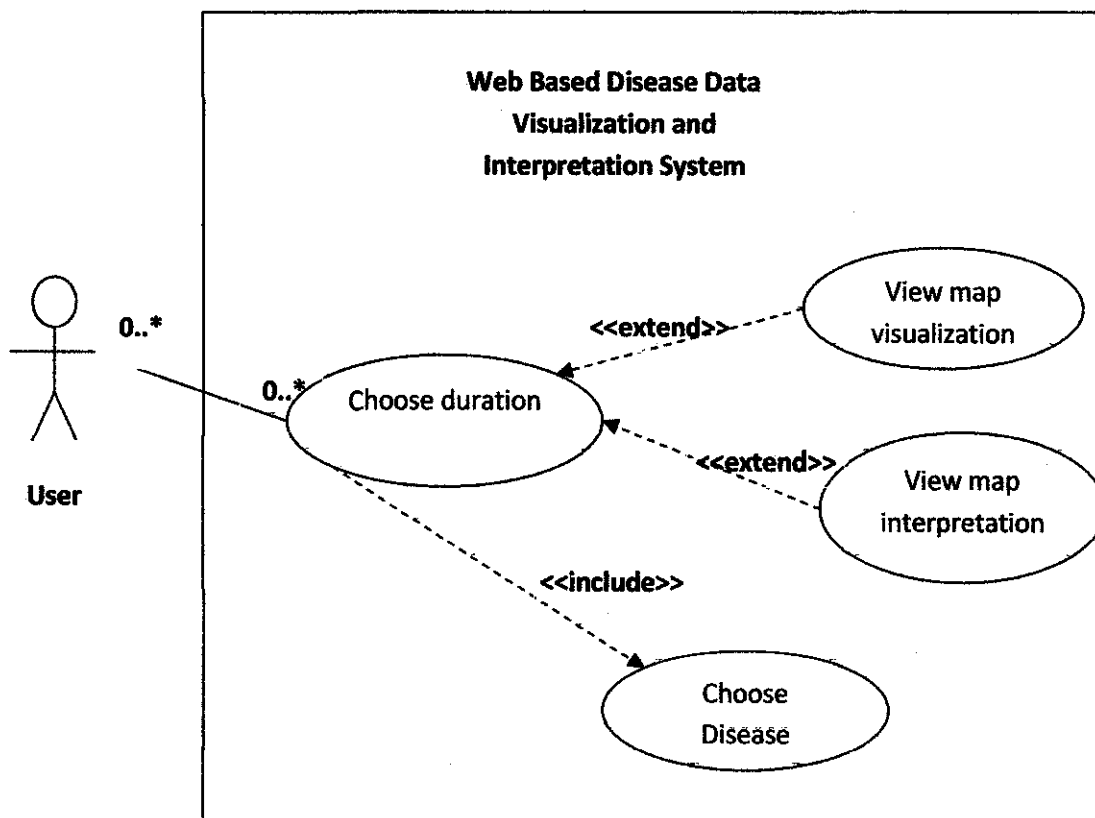


Figure 7: Use Case Diagram

4.2 Activity Diagram

Figure 8 below shows the activity diagram of the system. First, the user will choose the disease that he or she is interested to view. Then, the user needs to choose the area and duration for both the disease and disease related data that the user wants to view. The map will be generated based on users' selection. The data of the disease and weather will be compared to the value of the benchmark. For example, if the benchmark indicates that the disease cases above 5 in 20km² is considered high risk, then the interpretation will display that the area is a high risk area.

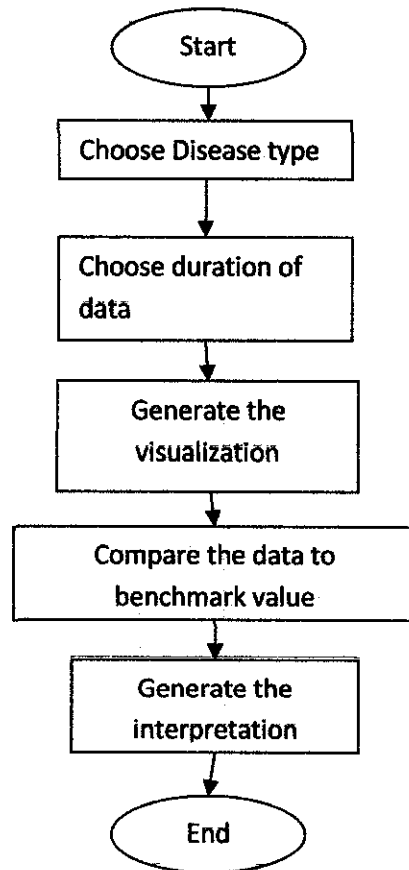
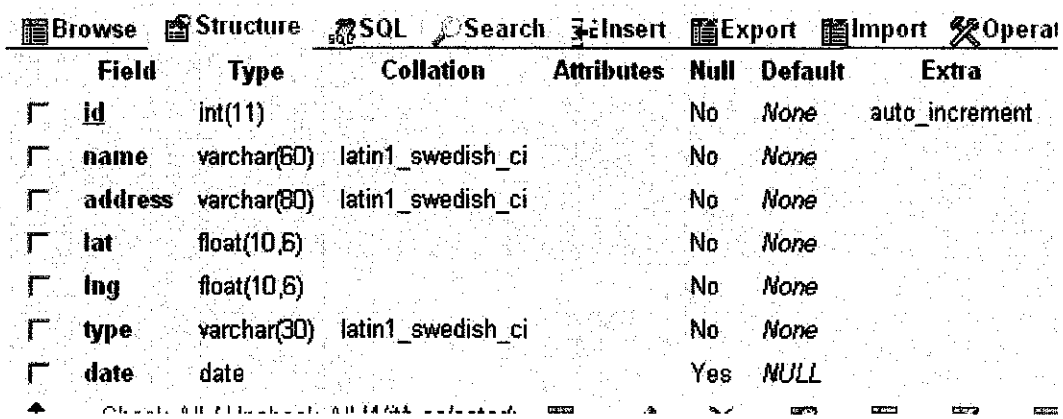


Figure 8: Activity Diagram

4.3 Database Design

4.3.1 Tables

The database consists of the markers table, a table that store disease types, disease date, disease address, and also longitude and langtitude of the address. Below is the table structure as shown in PhpMyadmin. Each year will be having a different table so that it is easier to show the disease data based on years.



Field	Type	Collation	Attributes	Null	Default	Extra
<input type="checkbox"/> id	int(11)			No	None	auto_increment
<input type="checkbox"/> name	varchar(60)	latin1_swedish_ci		No	None	
<input type="checkbox"/> address	varchar(80)	latin1_swedish_ci		No	None	
<input type="checkbox"/> lat	float(10,6)			No	None	
<input type="checkbox"/> lng	float(10,6)			No	None	
<input type="checkbox"/> type	varchar(30)	latin1_swedish_ci		No	None	
<input type="checkbox"/> date	date			Yes	NULL	

Figure 9 markers table (about disease)

4.3.2 Populate the data

Below is the screen shot of the raw Tuberculosis data sent by Mr. Deva:

A	B	C
Bil.	Kediaman Semasa (Seperti Dalam CDCIS)	Tarikh rawatan dimulakan
1	LOT 4228 KAMPUNG TERSUSUN KAMPUNG PULAI GOPENG	2009-10-29
2	NO. 39 PSRN. SARI 15 TAMAN CEMPAKA IPOH	2009-08-20
3	NO. 28 LALUAN HULU BERCHAM 4 BANDAR BARU PUTRA IPOH	2009-11-11
4	B-11 TKT. 11 PANGSAPURI SRI KEPAYANG, IPOH	2009-10-06
5	43, PERSIARAN RISHAH 7, KAWASAN PERINDUSTRIAN SILIBIN	2009-11-16
6	C25 JLN BENUS STARPARK IPOH PERAK	2009-11-16
7	357 A KAMPUNG BARU MAMBANG DIAWAN KAMPAR	2009-11-25
8	NO 460 JALAN BIDOR BUNTING IPOH	2009-08-10
9	707, LALUAN SIMEE 1, KG SIMEE IPOH PERAK	2009-11-10
10	BLOK C-29, NO. D-16, LENGKOK BERCHAM TIMUR 4, TMN PAKATAN IP	2009-11-11
11	NO. H10, PERUMAHAN AWAM 2, LAWAN KUDA BARU	2009-12-24
12	NO. 4 PSRN WIRA JAYA BARAT 29 TAMAN DESA INDAH IPOH	2009-11-19
13	NO. 58 CEMPAKA SARI 20 TAMAN CEMPAKA IPOH	2009-12-15

Figure 10 Raw Tuberculosis Disease Data

The author also need to add longitude, langtitude and type column after date so that the table will have same column with the table in the database. This will enable the import to be done. The author will put zeroes for the value of longitude and langtitude in Excel file, and put TB for the type if the disease is tuberculosis and Dengue for the type if the disease is Dengue. Moreover, since the date in phpMyAdmin database is year-month-date, the author need to change the date format in Excel file to this format before import. The author saved this Excel file sent by Mr. Deva as a .csv file and imported it to the database. The author has to change the address above to longitude and langtitude so that it can be marked on the map. After running the Geocoding script, it is found that not all addresses are successfully geocoded. The figure below is the screenshot of the .csv file exported from the database after the geocoding script being run. Not all addresses are geocoded. As a result, the author appends Perak, Malaysia behind the addresses to get a more accurate result. The append is done by using Excel concatenate function.

A	B	C	D	E	F
1	NO. 74, JALAN TAWAS, BATU 19, TAMAN EHSAN, IPOH, PERAK, MALAYSIA	10/2/2007	4.65184	101.0972	TB
2	36 LANDASAN TASEK TIMUR 7 TAMAN CAHAYA BERCHAM IPOH, PERAK, MALAYSIA	12/19/2007	4.646719	101.1256	TB
3	71 BUKIT MERAH NEW VILLAGE, LAHAT, PERAK, MALAYSIA	12/20/2007	4.543161	101.0407	TB
4	LOT 38007 JALAN BESAR SPG. PULAI KG. KEPAYANG IPOH, PERAK, MALAYSIA	12/27/2007	0	0	TB
5	8G KAMPUNG JAMBU TANJUNG RAMBUTAN, IPOH, PERAK, MALAYSIA	12/27/2007	0	0	TB
6	A7 JALAN LASAM GREENTOWN, IPOH, PERAK, MALAYSIA	12/27/2007	4.597343	101.0942	TB
7	24 HALA PKLN TIMUR 22 TAMAN PASIR PUTEH SELATAN IPOH, PERAK, MALAYSIA	12/27/2007	0	0	TB
8	23 JALAN TAWAS BARU 3 TAMAN TASEK DAMAI IPOH, PERAK, MALAYSIA	12/27/2007	0	0	TB
9	244, TAMAN SENTOSA KAMPAR, PERAK, MALAYSIA	7/4/2007	4.586623	101.0847	TB
10	53 PSRN. RAPAT BARU 20 TMN. LAPANGAN PERMAI, IPOH, PERAK, MALAYSIA	12/8/2006	0	0	TB
11	5589 KG. AIR RAPAT TG. TUALANG, PERAK, MALAYSIA	12/26/2006	0	0	TB
12	52 KG. BT. 9 TAMBAHAN JALAN JELAPANG, PERAK, MALAYSIA	12/17/2006	0	0	TB
13	56 JALAN LEONG LOH BOH, PERAK, MALAYSIA	1/4/2007	0	0	TB
14	56 JALAN LEONG LAH BOH, PERAK, MALAYSIA	1/4/2007	0	0	TB
15	NO. 63 KG. SINDU, KG. KEPAYANG, PERAK, MALAYSIA	1/5/2007	0	0	TB
16	NO.2 JALAN PEKELILING TASEK, PERAK, MALAYSIA	1/3/2007	4.606427	101.117	TB
17	3 BULATAN CHERRY TAMAN CHERRY, PERAK, MALAYSIA	1/7/2007	4.607386	101.0747	TB
18	NO. 37 HALA KLEDANG EMAS 6, TAMAN KLEDANG EMAS, IPOH, PERAK, MALAYSIA	1/9/2007	4.546847	101.074	TB
19	BLOK B, 30-4 KEM PGA ULU KINTA, PERAK, MALAYSIA	1/9/2007	35.11926	-95.2377	TB
20	D/A PENJARA BATU GAJAH, PERAK, MALAYSIA	1/9/2007	0	0	TB
21	40 PRSN ORKID 2, DESA CHANGKAT, BT GAJAH, PERAK, MALAYSIA	1/16/2007	0	0	TB
22	KG CHANGKAT LEGONG GOPENG, PERAK, MALAYSIA	1/16/2007	0	0	TB
23	57713 KG TASEK PERMAI, JLN GOPENG, BT GAJAH, PERAK, MALAYSIA	1/13/2007	0	0	TB
24	NO. 35 3552 KG TERSUSUN TANAH HITAM CHEMOR, PERAK, MALAYSIA	1/17/2007	0	0	TB
25	B-11 TKT. 11 PANGSAPURI SRI KEPAYANG, IPOH, PERAK, MALAYSIA	1/18/2007	0	0	TB
26	NO. 16 ULU KAMPAR, GOPENG, PERAK, MALAYSIA	1/11/2007	4.584835	101.1161	TB
27	NO. 16 ULU KAMPAR, PERAK, MALAYSIA	1/16/2007	0	0	TB

Figure 11 Append the state and country

After this, the author save this excel file in .csv extension. Then the author truncate the table in database and import this csv file. Below is the screen shot of the table before the Geocoding script being run.

Server: localhost Database: clean Table: markers

Browse Structure SQL Search Insert Export Import Operations Empty Drop

Showing rows 0 - 29 (137 total, Query took 0.0010 sec)

```

SELECT *
FROM 'markers'
LIMIT 0 , 20

```

Profiling [Edit] [Explain SQL]

Show 30 row(s) starting from record # 30

in horizontal mode and repeat headers after 100 cells

Sort by key: None

+ Options

	id	address	Date	lat	lng	Type
<input type="checkbox"/>	1	NO. 74, JALAN TAWAS, BATU 19, TAMAN EHSAN, IPOH ,P...	2007-10-02	0.000000	0.000000	TB
<input type="checkbox"/>	2	36 LANDASAN TASEK TIMUR 7 TAMAN CAHAYA BERCHAM IPO...	2007-12-19	0.000000	0.000000	TB
<input type="checkbox"/>	3	71 BUKIT MERAH NEW VILLAGE, LAHAT ,PERAK, MALAYSIA	2007-12-20	0.000000	0.000000	TB
<input type="checkbox"/>	4	LOT 38007 JALAN BESAR SPG. PULAI KG. KEPAYANG IPOH...	2007-12-27	0.000000	0.000000	TB
<input type="checkbox"/>	5	8G KAMPUNG JAMBU TANJUNG RAMBUTAN, IPOH ,PERAK, MA...	2007-12-27	0.000000	0.000000	TB
<input type="checkbox"/>	6	A7 JALAN LASAM GREENTOWN, IPOH ,PERAK, MALAYSIA	2007-12-27	0.000000	0.000000	TB
<input type="checkbox"/>	7	24 HALA PKLN TIMUR 22 TAMAN PASIR PUTEH SELATAN IP...	2007-12-27	0.000000	0.000000	TB
<input type="checkbox"/>	8	23 JALAN TAWAS BARU 3 TAMAN TASEK DAMAI IPOH ,PERA...	2007-12-27	0.000000	0.000000	TB
<input type="checkbox"/>	9	244, TAMAN SENTOSA KAMPAR ,PERAK, MALAYSIA	2007-07-04	0.000000	0.000000	TB
<input type="checkbox"/>	10	53 PSRN. RAPAT BARU 20 TMN. LAPANGAN PERMAI, IPOH ...	2006-12-08	0.000000	0.000000	TB
<input type="checkbox"/>	11	5589 KG. AIR RAPAT TG. TUALANG ,PERAK, MALAYSIA	2006-12-26	0.000000	0.000000	TB
<input type="checkbox"/>	12	52 KG. BT. 9 TAMBAHAN JALAN JELAPANG ,PERAK, MALAY...	2006-12-17	0.000000	0.000000	TB
<input type="checkbox"/>	13	56 JALAN LEONG LOH BOH ,PERAK, MALAYSIA	2007-01-04	0.000000	0.000000	TB
<input type="checkbox"/>	14	56 JALAN LEONG LAH BOH ,PERAK, MALAYSIA	2007-01-04	0.000000	0.000000	TB
<input type="checkbox"/>	15	NO. 63 KG. SINDU, KG. KEPAYANG ,PERAK, MALAYSIA	2007-01-05	0.000000	0.000000	TB

Soft32Downloader-for...exe markers TB 2007 until .csv thuy.csv markers.xls

localhost / ... XAMPP C... 2 Windo... Celcom Br... Downloadi... after geoco...

Figure 12 Before run geocoding

After the author runs the geocoding php file, the author check the database again. This time it shows that the longitude and longitude already populated.

Showing rows 0 - 29 (137 total, Query took 0.0006 sec)

```

SELECT *
FROM markers
LIMIT 0, 30

```

Profiling [Edit] [Explain SQL] [Create PHP Co

Show 30 row(s) starting from record # 30

in horizontal mode and repeat headers after 100 cells

Sort by key: None

+ Options

	id	address	Date	lat	lng	Type
<input type="checkbox"/>	1	NO. 74, JALAN TAWAS, BATU 19, TAMAN EHSAN, IPOH ,P...	2007-10-02	4.651840	101.097153	TB
<input type="checkbox"/>	2	36 LANDASAN TASEK TIMUR 7 TAMAN CAHAYA BERCHAM IPO...	2007-12-19	4.646719	101.125633	TB
<input type="checkbox"/>	3	71 BUKIT MERAH NEW VILLAGE, LAHAT ,PERAK, MALAYSIA	2007-12-20	4.543161	101.040733	TB
<input type="checkbox"/>	4	LOT 38007 JALAN BESAR SPG. PULAI KG. KEPAYANG IPOH...	2007-12-27	4.611750	101.113503	TB
<input type="checkbox"/>	5	8G KAMPUNG JAMBU TANJUNG RAMBUTAN, IPOH ,PERAK, MA...	2007-12-27	4.611750	101.113503	TB
<input type="checkbox"/>	6	A7 JALAN LASAM GREENTOWN, IPOH ,PERAK, MALAYSIA	2007-12-27	4.597343	101.094177	TB
<input type="checkbox"/>	7	24 HALA PKLN TIMUR 22 TAMAN PASIR PUTEH SELATAN IP...	2007-12-27	4.611750	101.113503	TB
<input type="checkbox"/>	8	23 JALAN TAWAS BARU 3 TAMAN TASEK DAMAI IPOH ,PERA...	2007-12-27	4.660783	101.110756	TB
<input type="checkbox"/>	9	244, TAMAN SENTOSA KAMPAR ,PERAK, MALAYSIA	2007-07-04	4.586623	101.084656	TB
<input type="checkbox"/>	10	53 PSRN. RAPAT BARU 20 TMN. LAPANGAN PERMAI, IPOH ...	2006-12-08	4.611750	101.113503	TB
<input type="checkbox"/>	11	5589 KG. AIR RAPAT TG. TUALANG ,PERAK, MALAYSIA	2006-12-26	4.607294	100.800003	TB
<input type="checkbox"/>	12	52 KG. BT. 9 TAMBAHAN JALAN JELAPANG ,PERAK, MALAY...	2006-12-17	4.626612	101.080570	TB
<input type="checkbox"/>	13	56 JALAN LEONG LOH BOH ,PERAK, MALAYSIA	2007-01-04	4.807294	100.800003	TB
<input type="checkbox"/>	14	56 JALAN LEONG LAH BOH ,PERAK, MALAYSIA	2007-01-04	4.807294	100.800003	TB
<input type="checkbox"/>	15	NO. 63 KG. SINDU, KG. KEPAYANG ,PERAK, MALAYSIA	2007-01-05	4.807294	100.800003	TB

markers TB 2007 until.csv thuy.csv markers.xls Show a

localhost /... XAMPP C... Macromedi... My Comp... Calcom Br... Downloadi...

Figure 13 After run geocoding

4.4 Display Map

In order to display the map, we need to create markers for the each disease cases. When clicked on the markers, the name of the markers which is the full address of the disease case will be shown. The author use Google Maps API to create markers and to show the map. Note that the markers below are red and blue balons.

4.4.1 Display Map for different years

Below are the results for Tuberculosis case for Daerah Kinta from year 2007 to year 2010.

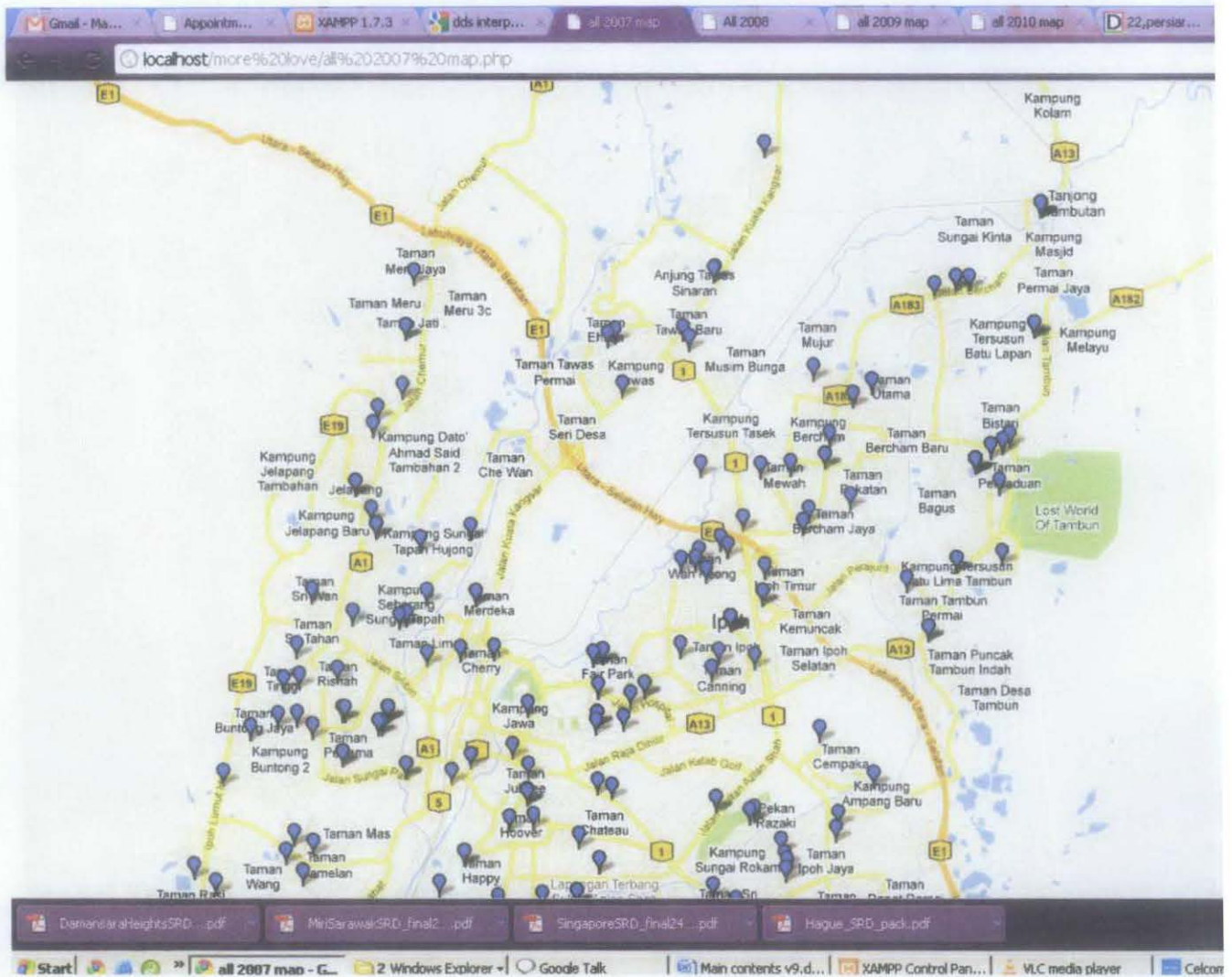


Figure 14 Tuberculosis for year 2007

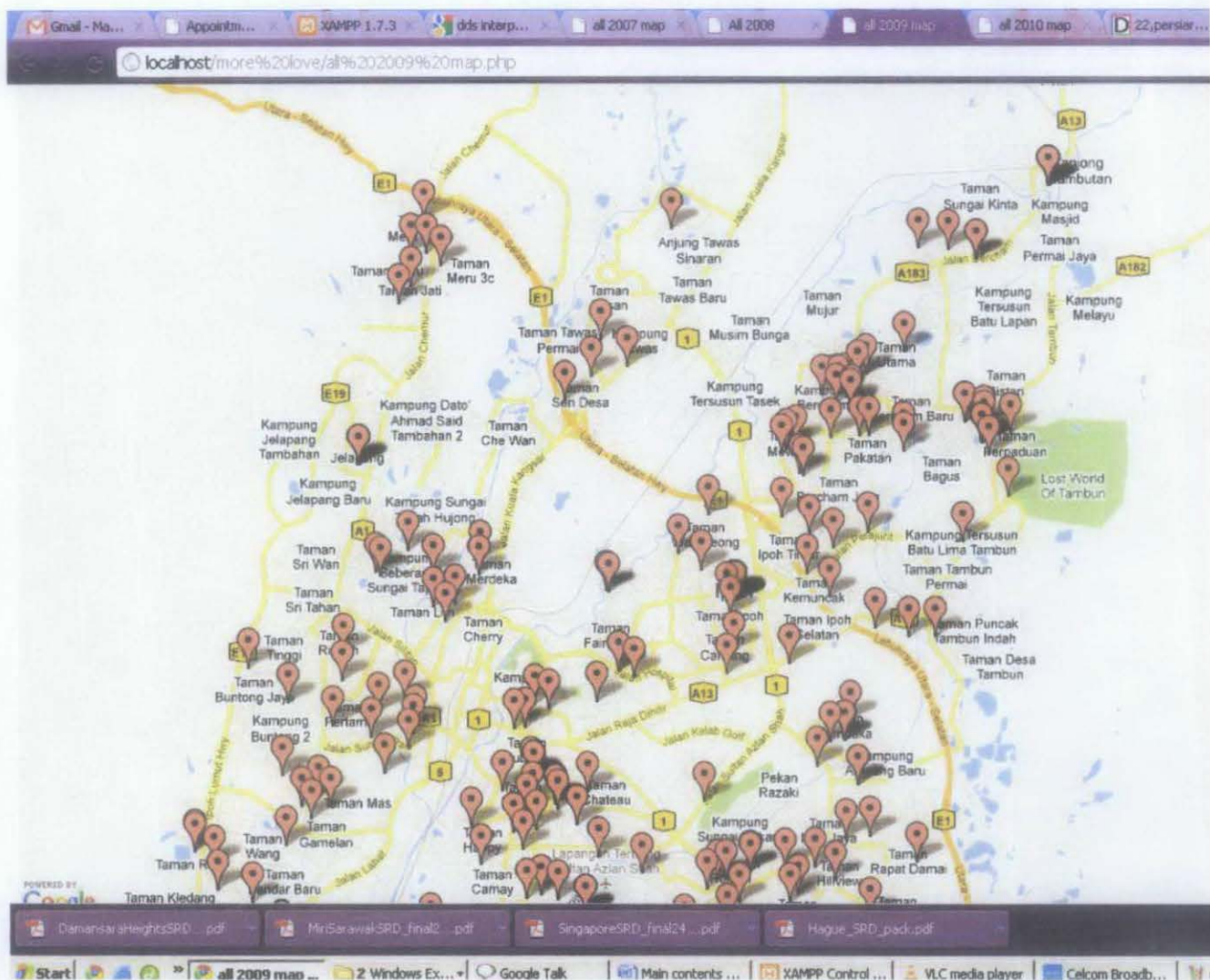


Figure 16 Tuberculosis for year 2009

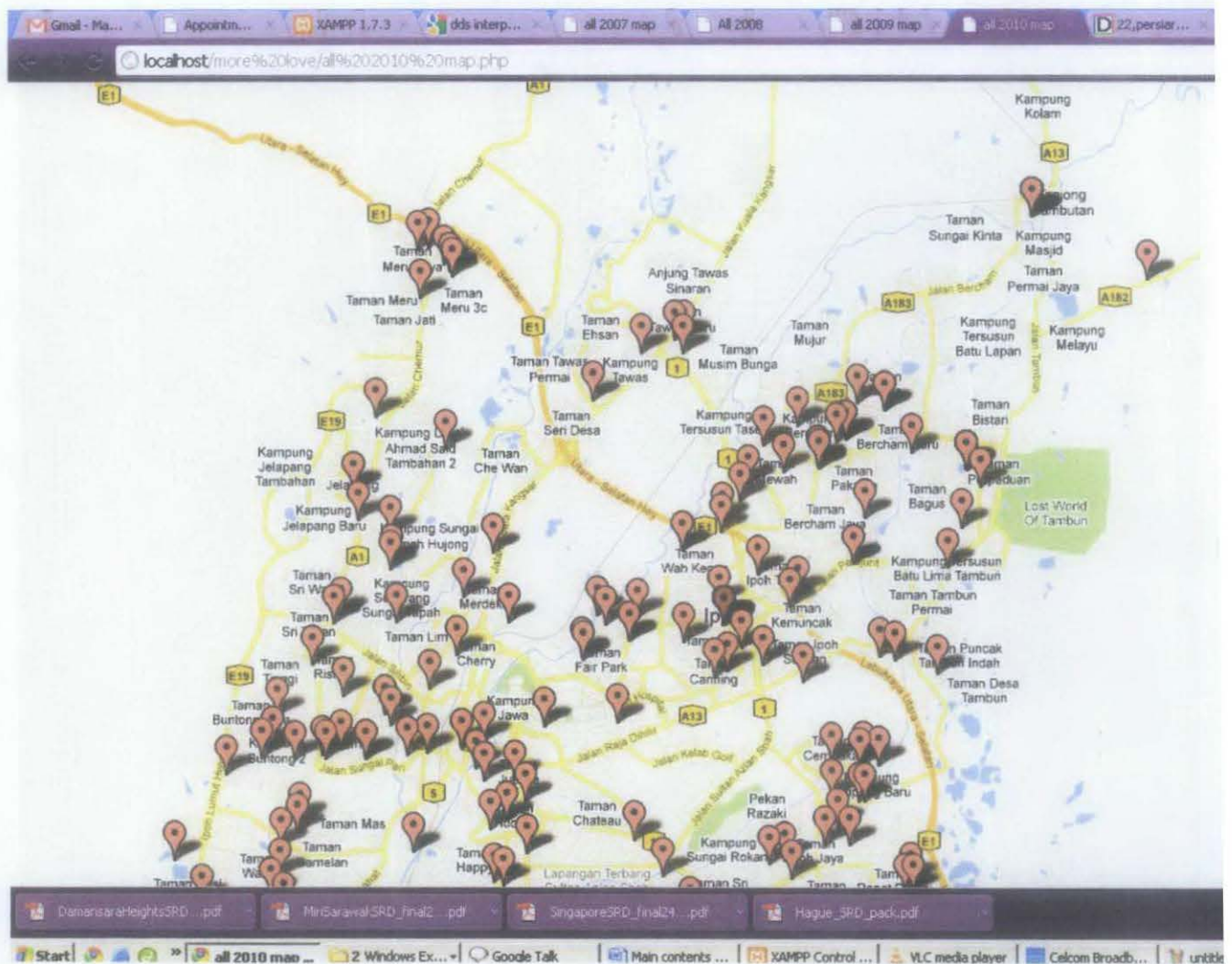
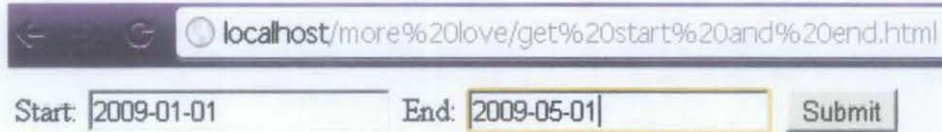


Figure 17 Tuberculosis for year 2010

Note that the markers used is different where the year 2007 and 2008 are using blue markers and year 2009 and 2010 is using red markers. The markers used can be specified in the javascript code to show the markers. Thus, it is possible to show disease data from different year on the same map where the disease data from different years are represented by different color's markers.

4.4.2 Query the map

Other than display the diseases for one whole year, the user can also choose to display the disease cases happened on a certain period of time only. They can do so by entering the start date and end date of the period of disease cases where they want to view:



The image shows a browser window with the address bar containing the URL `localhost/more%20love/get%20start%20and%20end.html`. Below the address bar, there is a form with two text input fields. The first field is labeled "Start:" and contains the date "2009-01-01". The second field is labeled "End:" and contains the date "2009-05-01". To the right of these fields is a "Submit" button.

Figure 18 Choose Start and end date

After the user chooses the start and end date for the period, the user will be redirected to another page where the map will only show the markers of disease cases happened within the period the users entered.

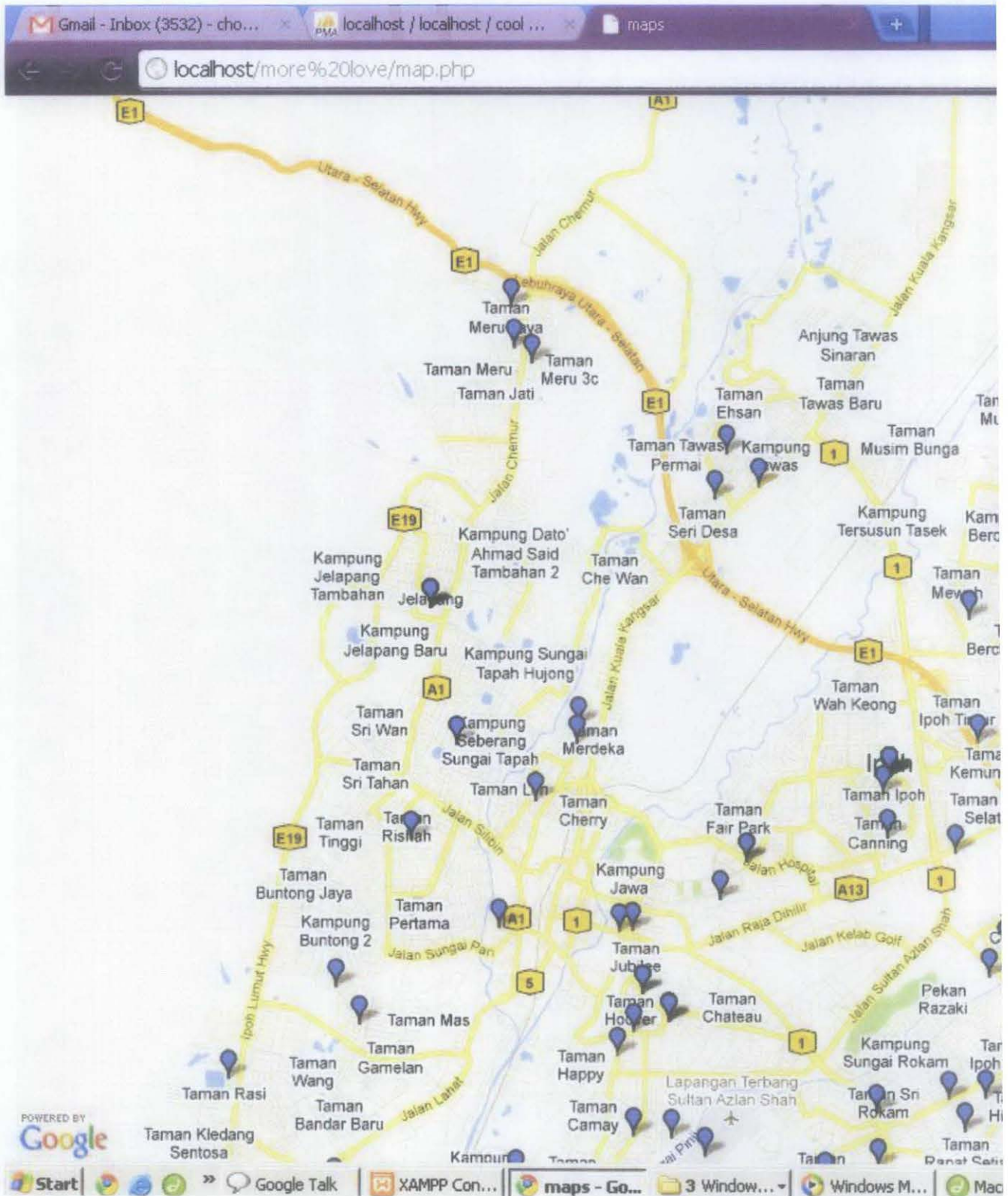
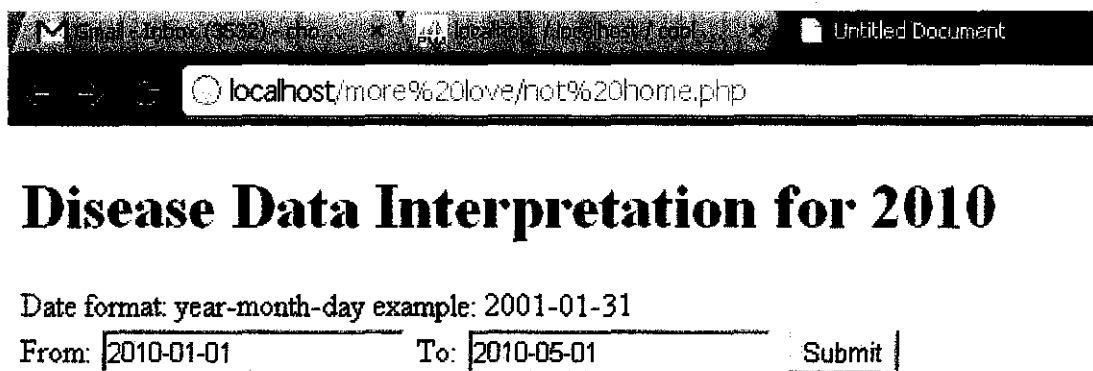


Figure 19 Map that shows only diseases within user queried date

4.4.3 Show Interpretation

The interpretation are generated based on the count(*) function in sql where the number of cases within the period chosen by the users will be calculated and will be compared against the standard provided by Jabatan Kesihatan Daerah Kinta to determine the disease level(high, medium, low). The standard will be different for different type of disease.

The user have to open the page for interpretation and enter the start and end date.



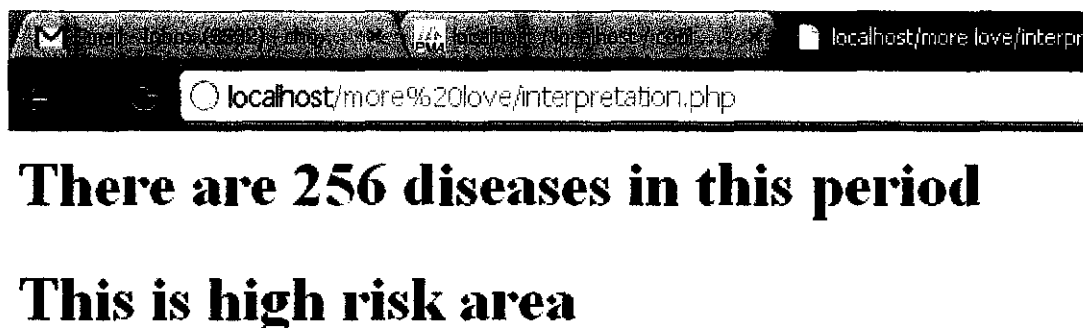
localhost/more%20love/not%20home.php

Disease Data Interpretation for 2010

Date format: year-month-day example: 2001-01-31

From: To:

Figure 20 Choose start and end date for Interpretation



localhost/more%20love/interpretation.php

There are 256 diseases in this period

This is high risk area

Figure 21 Interpretation

4.5 Discussion

According to Mr. Deva, the head of Data Surveillance Unit of Jabatan Kesihatan Daerah Kinta, this system is very useful for them to track the outbreak of the communicable diseases in a timely basis. Mr. Deva give the example where there used to be high density of diarrhea vomit cases in one area that is related to communicable diseases in one area few years ago. The health officer had investigated the matter and realized the broken of an underground water pipe which allow microorganisms to enter is the cause. They had fixed the problem and few years later, if there is again high density of diarrhea and vomit in the same area, they can spot the problem immediately if system. If not, it is very hard for them to link the past to current scenario since it is very hard to visualize the address and density of the disease case by just looking at the disease record which contains only words.

Mr. Deva suggested the author to create a nice interface for this system so that the health officer with no knowledge of how to use phpMyAdmin to insert the data to the database can use the system without effort.

On the other hand, the author also suggested the health institution to obtain each element of the address from the patient separately because it will make the job of geocoding easier. This is because the website that provide batch geocoding other than google geocoding require the elements of the address such as Street, state and country to be entered separately. If the address given by the health institution is the full address, the batch geocoding cannot be done and we have to use google geocoding to do the geocoding.

CHAPTER 5

CONCLUSION

This report provides an overview of the Disease Data Visualization and Interpretation System, with the aim to ease the analysis of disease data by visualizing them on map. The data will be visualized on the map showing the location where the diseases happen. A short paragraph will also be generated to interpret the data that are being visualized. Prototyping methodology is chosen because it allows the generation of mockups for testing and user interactions. From the literature review of the Web GIS based Information Visualization for Infectious Disease prevention, the author found that it is very useful to have the disease data visualize on the map. This is due to the fact that such application will be helpful in tracking the infectious diseases' spreading tendency, in building up isolation for the infectious disease and also in establishing the best transportation line for the personnel and the equipment supply in the infectious region. As a result, the author decided to build a similar application which focuses on Daerah Kinta of Perak state of Malaysia. A paragraph is added to give an impartial explanation of the data being visualized. Based on the comments of the potential users, the system can be enhanced by connecting the application to the local hospital database to obtained direct update about the disease data. With this, the users can just access this website to get the most up to date disease information and be able to plan their journey more effectively to avoid the high risk area. Last but not least, the author had learnt a lot from this project and will continue to be committed in doing this project to ensure the success of this project so that it can provide convenience to the end users.

REFERENCES

1. WHO Western Pacific Region - Home. (n.d.). *World Health Organization Western Pacific Region*. Retrieved August 25, 2010, from www.wpro.who.int/internet/files/eha/toolkit/web_bakup/country_profiles.htm
2. Merler, S., Jurman, G., Furlanello, C., Rizzo, C., Bella, A., Massari, M., et al. (2006). Strategies for containing an influenza pandemic: The case of Italy. *Bio-Inspired Models of Network, Information and Computing Systems, 2006*. 1st, 1-7.
3. Richards, G. (2009). Don't panic! - [control medical]. *Engineering & Technology, 4*(10), 45-47.
4. Google Flu Trends | How does this work?. (n.d.). *Google.org - Google Technology-Driven Philanthropy*. Retrieved October 12, 2010, from <http://www.google.org/flutrends/about/how.html>
5. Google Insights for Search. (n.d.). *Google*. Retrieved October 12, 2010, from <http://www.google.com/insights/search/?hl=en-GB#q=dengue&geo=MY&date=1%2F2009%2012m%2C1%2F2010%2012m&cmpt=date>
6. Perry. (n.d.). Asthma and Humidity. *Health Guidance - Free Health Articles*. Retrieved October 12, 2010, from <http://www.healthguidance.org/entry/10912/1/Asthma-and-Humidity.html>
7. Palaniappan, S., & Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques. *Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on*, 108-115.
8. Hailiang Jin, & Liu, H. (2009). Research on visualization techniques in data mining. *Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on*, 1-3.
9. Kantardzic, M.; , "Concepts, Models, Methods, and Algorithms," *Data Mining* , pp.360
10. Tsui, K. L., Goldsman, D., Jiang, W., & Wong, S. Y. (2010). Recent research in public health surveillance and health management. *Prognostics and Health Management Conference, 2010. PHM '10*. 1-7.

11. Omar, S. (2007). *Geographical Information System for Dengue Management System*. Tronoh: Universiti Teknologi PETRONAS, 1-9.
12. Syahman, A. M. (2006). *GIS for Dengue Epidemic Management for Ipoh/ Batu Gajah/ TBD*. Tronoh: Universiti Teknologi PETRONAS, 1-10
13. Wenhua Zeng, Xia Cui, Xiangnan Liu, Haishan Cui, & Ping Wang. (2006). Remote sensing and GIS for identifying and monitoring the environmental factors associated with vector-borne disease: An overview. *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on*, 1443-1444
14. Tan Kah Bee, Koh Hock Lye, & Teh Su Yean. (2009). Modeling dengue fever subject to temperature change. *Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09. Sixth International Conference on*, 5 61-65.
15. Definition of visualize. (n.d.). *Yahoo! Education*. Retrieved October 13, 2010, from education.yahoo.com/reference/dictionary/entry/visualize
16. Data Visualization. (n.d.). *Department of Computer Science - Welcome*. Retrieved October 12, 2010, from <http://web.cs.wpi.edu/~matt/courses/cs563/talks/datavis.html>
17. Friedman, V. (n.d.). Data Visualization and Infographics - Smashing Magazine. *Smashing Magazine*. Retrieved October 12, 2010, from <http://www.smashingmagazine.com/2008/01/14/monday-inspiration-data-visualization-and-infographics>
18. Fayyad, U., Grinstein, G., & Wierse, A. (2001). *Information Visualization in Data Mining and Knowledge Discovery (The Morgan Kaufmann Series in Data Management Systems)* (1st ed.). San Francisco: Morgan Kaufmann.
19. 5 Best Data Visualization Projects of the Year. (n.d.). *FlowingData | Data Visualization, Infographics, and Statistics*. Retrieved October 13, 2010, from <http://flowingdata.com/2008/12/19/5-best-data-visualization-projects-of-the-year/>
20. Crazyegg. (n.d.). *crazyegg- build heatmap and track clicks*. Retrieved October 13, 2010, from <http://www.crazyegg.com/>
21. Xiaolin Lu. (2009). Web GIS based information visualization for infectious disease prevention. *Intelligent Information Technology Application, 2009. IITA 2009. Third International Symposium on*, 1 148-151.

22. PD-Downloads : Pejabat Kesihatan Wilayah Persekutuan Putrajaya. (n.d.). *Kami Sedia Membantu : Pejabat Kesihatan Wilayah Persekutuan Putrajaya*. Retrieved October 13, 2010, from <http://pkpj.moh.gov.my/modules/PDdownloads/viewcat.php?cid=2>}
23. Google Maps Javascript API V3 Services - Google Maps JavaScript API V3 - Google Code. (n.d.). *Google Code*. Retrieved October 20, 2010, from <http://code.google.com/apis/maps>

APPENDICES

APPENDIX 1

KEY MILESTONES

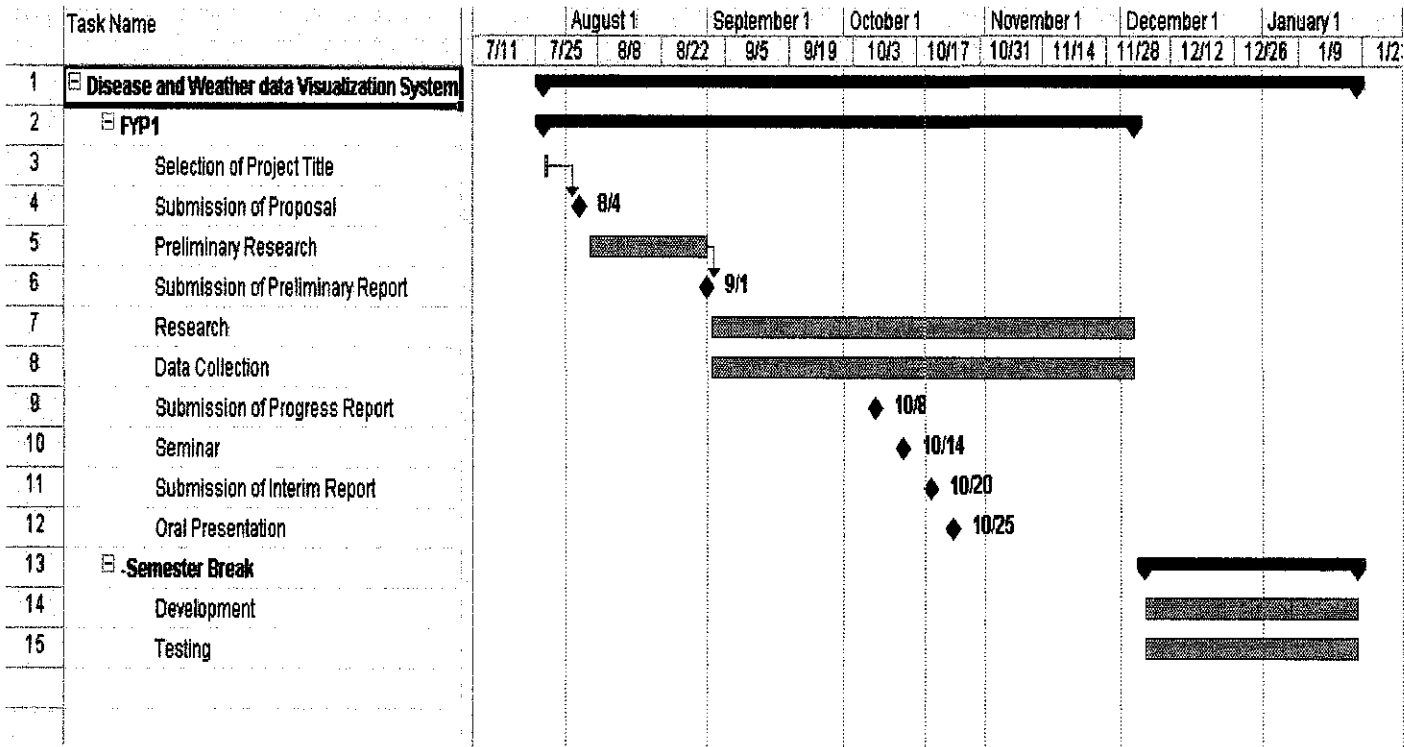
Submission of proposal	4	August	2010
Submission of preliminary report	18	September	2010
Submission of progress report	8	October	2010
Seminar	14	October	2010
Submission of interim report	20	October	2010
Oral presentation	25	October	2010

FYP2

Progress Report	16	March	2011
Pre EDX	30	March	2011
Dissertation	11	April	2011
Viva	20	April	2011

APPENDIX 2

PROJECT GANTT CHART



Gantt chart for FYP 2

FYP 2	Jan	Feb	Mar	Apr	May
Development and testing					
technical report					
final dissertation					
pre edx					
viva					



UNIVERSITI
TEKNOLOGI
PETRONAS

Dear Respondent,

Permission to Collect Disease Data for the Research Purpose

Ms. Chow Cheah Hui is a Final Year student in Computer and Information Sciences Department at Universiti Teknologi PETRONAS. She is conducting a research on the relationship between occurrence of diseases in Malaysia and other elements such as weather or the Google search trend of the disease related keywords. The research aims to find out the relationship between the said elements and thus predict the future outbreak of the disease.

Your cooperation to provide the related data would be very helpful for her research.

Thank you.

Yours faithfully,

Ms. Goh Kim Nee

Lecturer.

Department of Computer and Information Sciences

Universiti Teknologi PETRONAS

Bandar Seri Iskandar

P.O.Box 31750, Tronoh.

Perak Darul Ridzuan, Malaysia.

E-mail: gohkimnee@petronas.com.my

Chow Cheah Hui
ICT Programme,
Village 5 G,
Universiti Teknologi PETRONAS,
31750 Tronoh, Perak Darul Ridzuan.
Malaysia.
Tel : +60195265184
Email : chowehuir2006@gmail.com

24th February 2011

Pegawai Kesihatan Daerah,
Pejabat Kesihatan Daerah Kinta,
JALAN AMAN,
31000 BATU GAJAH,
PERAK,
MALAYSIA
Phone: 05-3652062 / 2072
Fax : 05-3668073

Dear Sir/ Madam:

Subject: Request for communicable disease data

I, Chow Cheah Hui, is a final year student of Universiti Teknologi PETRONAS. I am currently doing a project named Web based Disease and Weather Data Visualization and Interpretation System. I would like to request for some data that are significant for this project which can benefit the nation by visualizing the data on the map which ease the health officer's job to analyze the data .

The project intends to collect the information about the communicable diseases data and the weather data and place the data on the map where the disease happened. Then a paragraph of interpretation will be generated based on the condition of the data. Research about the disease will be done to set the standard of interpretation. The project will visualize the disease data and the weather data and can help the health centre officer to analyze the disease. The public can also use the data to learn more about the behavior of the disease.

I would like to request for the data of Tuberculosis (TB) and Dengue for my research. **The data required is only the date and address for each TB and Dengue case.** The period of time is from 2005 up to 2011. It is more desirable to get the soft copy of the data if any because this will speed up the data processing process.

Please feel free to call me at 0195265184 or email me at chowehuir2006@gmail.com if you have any question.

Respectfully Yours,



(Chow Cheah Hui)
Information & Communication Technology
Final Year Final Semester
Universiti Teknologi PETRONAS