MASHUP FOR INTERNSHIP PLACEMENT (MIP)

by

Norfarahin Adiba Binti Abd Kadir

10918

Dissertation submitted in partial fulfillment of

The requirements for the

Bachelor of Technology (Hons)

(Information and Communication Technology)

JANUARY 2011

Universiti Teknologi PETRONAS

Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Mashup for Internship Placement (MIP)

by

Norfarahin Adiba Binti Abd Kadir

A project dissertation submitted to the
Computer Information Science Department
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
BACHELOR OF TECHNOLOGY (Hons)
(Information and Communication Technology)

Approved by,

____
(Ms. Amy Foong Oi Mean)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
APRIL 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 here in have not been undertaken or done by unspecified sources or persons.

NORFARAHIN ADIBA BINTI ABD KADIR

ABSTRACT

Mashup, a web application that integrates data from multiple data sources or APIs to provide unique services. In this project Mashup for Internship Placement (MIP), involves solving multiple problems such as extracting data from multiple sources, cleaning it, and integrate it together. In the case study presented in this paper, Mashup has been applied as one of the pattern for knowledge sharing. The problem statement for this research paper is there is difficulties for student to decide which internship offer to accept because they lack of knowledge about the company and the job description. The objective of the research are to investigate Mashup pattern for knowledge sharing system and to develop a prototype system that incorporate features that allow user to view the location and information about the company which offer the internship placement. The methodology used to create Mashup system is Prototyping based methodology and Yahoo Pipes as the platform. Data from Google Maps is extract using Fetch Feed widget in Yahoo Pipes and integrate the feeds with Yahoo Pipes to show result on interactive map rather than traditional list. MIP contains information about company that offers internship placement with the location view on the map. This Mashup also can help student to be focus on the decision making for their internship placement as they may get all information gathered in one application.

ACKNOWLEDGMENTS

This work has benefited from the input of many people. The author would like to thank Ms Amy Foong Oi Mean for her guide throughout the project completion and her valuable feedback on the project. Particular thanks should go to Mr Rabiu Ibrahim for his guide and contribution towards the project, author's friends, Wan Aathena Wan Ahmad Marzuki, Anis Akmal Binti Anuar, Nurul Ain Binti Md Yazid, Nik Nor Ernina Binti Nik Ab Rahman, Nur Hidayah Binti Fadzil and Siti Ain Nurena Binti Mohd Nasir for their support and motivation to complete this project.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Recently, the new kind of data integration and information application is emergent up all across the Internet. In the past, knowledge sharing system is not much exciting as compared to these days with lots of interactive applications that draw upon contents and integrate data from multiple web sources to provide unique and innovative services. That is what Mashup is doing.

This system integrates information about the companies, for example location of the company on the maps, with the address and extra information about the company will be gathered on the map in the application. This Mashup is built as a pipeline linking the search logs by the user and integrating it with Google maps from admin database. The results then are displayed on a map widget using a marker which will highlight the number of companies based on the categories and the location of it on the maps.

This first part of this paper, define the existing problem of the current system and the objective as the solution for all the problems and the scope of study. Second part discusses the literature review based on other project that has similarities, compares and make some improvement to the system. The third part of this paper highlights the methodologies to be used to develop this system. Focuses on techniques and design used to build Mashup. Forth part discuss on the result and discussion from the progress of system development. Chapter five conclude everything and main findings from this project.

1.2 PROBLEM STATEMENT

1) Difficulties for student to decide which internship offer to accept because they lack of knowledge about the company and the job description.

1.3 OBJECTIVES

1) To investigate a Mashup pattern for knowledge sharing system.

2) To develop a prototype system that incorporate features that allow user to view the location and information about the company which offer the internship placement.

1.4 SCOPE OF STUDY

The scope of study in this project is to focuses on a simple Mashup application used to track the location of companies for industrial internship in interactive way which will be view on the map as the users nowadays tend to be more visual rather than just a list of it. This research will focus on the internship placement available in Selangor and Kuala Lumpur area. The Mashup will only focus on company that offer placement which is suit the course that UTP offer to their student which are the engineering industry, business industry, information technology industry and not to forget the oil and gas industry.

Using Yahoo Pipes as the platform to create Mashup for internship placement is also one of the scope to this research paper. As there exist lots of Mashup building tools, such as Openkapow (kapowsoftware.com), Dapper (www.dapper.net), IBM Mashup center (www-01.ibm.com), WaveMaker Studio (www.wavemaker.com) and Yahoo Pipes (pipes.yahoo.com) to name a few. Yahoo Pipes allow the developer to build Mashup without writing code. However, not having writing code to build Mashup does not always mean building it is easy [18]. There might be some problem to integrate different data from data from different multiple sources. The issue of data integration may arise if the pipes do not link properly as the result will show a Mashup with no map for example.

There are two problems with the widget approach as Yahoo Pipes using widget. First, the number of widgets increasing as Mashup increasing in the functionality. As the result, locating a widget to complete the Mashup function is difficult and time consuming. Second, there are no programming is required but the need to understand the concept of programming to fully utilize the widget is very important [18]. Even though there is some issue regarding the integration between data from different multiple sources, Yahoo Pipes give user to select and customize widget to perform complex operations.

Mashup for internship placement will help users using the information and shared knowledge faster and efficient as users can get all information in one application without need to search other related information in others website. All important information needed about the companies will be combined and tagged together and view it on the map.

CHAPTER 2

LITERATURE REVIEW

As clear stated in the introduction, Mashup is one of the interactive web application that integrate and combine data from other sources to create an innovative and creative application on the maps. This application can be done by many different ways, using REST (Representational State Transfer) protocol, JSON (Java Script Object Notation), RSS (Really Simple Syndication) feed or existing widget.

If a web site uses data or functionality from another web site and combines it in an application, it's a Mashup. The application can access the data or functionality in various ways. It can use formal Representational State Transfer (REST)-based APIs provided by the other site. Or it can do some informal screen scraping, in which it extracts data from the displayed output of a program on another site. Or it can access an RSS feed or use a widget provided by another site.[4]

2.1 CHALLENGES IN CREATING MASHUP

Creating a Mashup, a web application that integrates data from various web sources to provide single services visualizes on the map solves multiple challenges, such as extracting data from multiple web sources, cleaning it and integrate it together [3]. This research paper highlights some of the common integration challenges involving in building the Mashup. These can be summarized as follows.

Data Extraction. Information relevant for the Mashup is pushed from a set of data sources. In this case study, the search button must be monitored for changes made by the user input of search for company name or location and each new entry must extract information of company location on the map. The fetch widget in Yahoo Pipes will extract information from Google Maps where a dataset is created for list of hosted company that offer internship placement. To build a list of company in Google Maps, the database use Google Maps application to extract the location as it will make the database more accurate. Widget in Yahoo Pipes extract the RSS feed from Google Maps.

Data Cleaning. Different data sources may present data in different format, so cleaning is required to fix misspelling and transform extracted data into an appropriate format [3]. In Yahoo Pipes data cleaning is done by putting a filter to the information that has been extracted. When the information if filter so the Mashup will give result base on what user input. If it is not filter, then the information might not be accurate according to what user request.

Data Integration. From definition, in a Mashup data is combine from multiple web source into one unique application. This system integrates data for location of companies and number of student doing internship there for certain period of time or based on the year. For example joining data extracted from user search with geocoding information stored in database. [2]

Service Heterogeneity. Access to data source may take place under different style : a subscription to an RSS feed, a traditional Web services invocation SOAP, the retrieval of the resources state a RESTful Web service [5], or by piping the result of UNIX command line. Regarding the example Mashup, the geocoding database usually can be accessed remotely through a web service interface [10, 11, 12] which may however put a daily cap to limit the number of messages exchanges. As more efficient alternative, a JDBC interface may available to directly query a locally installed copy of the database [13, 14, 15].

Maintainability. The long-term maintainability of a Mashup is directly correlated with the amount of change affecting the APIs of the composed Web services. No-longer maintained Mashups are very likely to break as the underlying APIs evolve independently, and the data sources integrated by the Mashups change their representation format. Having a clear model of the example, when running the Mashup example with different Web server, only the data extraction part of the pipeline needs to be updated in case the access log format has changed [2].

2.2 TYPICAL MASHUP ARCHITECTURE

As proposed by Duane Merrill in his paper entitle 'Mashups: The new breed of Web App' (Aug, 8th 2006) he discuss that a Mashup application sis architecturally comprised of three different participants that are logically and physically disjoint: API/content providers the Mashup site, and the clients Web browser.

- **The API/content provider**. These are the providers of the content being mashed. In this case study, the providers are Google and the SIIU department that provides the data. To facilitate data retrieval, providers often expose their content through Web-protocols such as REST, Web Services, and RSS/Atom. However many interesting potential data-sources so not yet conveniently expose APIs. Mashups that extract content form sites like Wikipedia, TV Guide, and virtually all government and public domain Web sites do so by a technique known as screen scraping.[6]

- **The mashup site**. This is where Mashup is hosted. Interestingly enough, just because this is where the Mashup logic resides, it is not necessarily where it is executed. On one hand, Mashups can be implemented similarly to traditional Web application using server-side dynamic content generation technologies like Java servlets, CGI, PH or ASP. For this project, Mashup for internship placement is hosted in Yahoo Pipes. Using Yahoo Pipes widget, the Mashup engine is created than after that can be embed any HTML or other website that may need to use this application.

  Alternatively, mashed content can be generated directly within the client's browser through client-side scripting or applets. This client-side logic is often the combination of code directly embedded in the Mashup's Web pages as well as scripting API libraries or applets. Mashup using this approach can be termed Rich Internet Applications (RIAs), meaning that they are very oriented towards the interactive user experience. The benefits of client-side mashing include less overhead on behalf of the Mashup server (data can be retrieve directly from the content provider) and a more seamless user-experience (pages can request updates for portions of their content without having to refresh entire page). The Google Maps API is intended for access through browser-side JavaScript and an example of client-side technology. [6].

- **The client's Web browser**. This is where the application is rendered graphically and where user interaction takes place. The Mashup engine or application can be embedded into client web browser by using RSS feed. They just need to copy the URL of the RSS and paste it on the browser address bar. The browser will show the result of Mashup created using Yahoo Pipes.

From a journal title “Intel Mash Maker : Join the Web” a research paper by Rob Ennals, Eric Brewer, Prasahant Gandhi all are Intel researcher, Minor Garofalakis a Yahoo researcher and Michael Shadle who are from Software Solutions Group of Intel Corporation, they write about Intel Mash Maker which is an interactive tool that tracks user activities and use the information for their current task.

“The Intel Mash Maker client is currently implemented as an extension to the FireFox web browser. Mash Maker adds a toolbar to the browser that shows buttons representing enhancements that Mash Maker believes the user might want to apply to the current page. An enhancement might combine the data in a new way. Mash maker is intended to be an integral part of the way the user use browses information, rather than being a special tool that a user uses when they want to create Mashups. In order to create Mashups from normal website, Mash Maker must first extract structured data from them. If the website does not provided RDF data, then the Mash Maker extracts structured data from the raw HTML using a community-maintained database of extractor, where each extractor describes how to extract structured data from a particular kind of website” [8].

2.3 SUN DATA MASHUP ENGINE

The Sun Data Mashup Engine is a JBI-compliant service engine that provides Enterprise Data Federation, or Mashup services, in multiple forms. It provides a single view of data from dissimilar sources, including static web pages, tabular data exposed as web services, relational databases, flat files, and so on. The Sun Data Mashup Engine joins, aggregates, and cleanses the source data, and then exposes the resulting data set in a unified view, providing real-time views of information for master data consumers.[16]

The Sun Data Mashup Engine provides the following features:

- Exposes data services for Web 2.0 Mashups and real time views for MDM clients.
- Supports JSON as well as Web Row Set.
- Supports both SOAP and REST invocation.
- Leverages various Binding Components (BCs) to access ERP/CRM systems. The XSLT Service Engine can be used for further transformation.
- Creates a unified view of data from disparate sources.
- Integrates information from a variety of heterogeneous sources, including relational databases, flat files,DCOM documents, spreadsheets, XML documents, HTML documents,
- RSS and Atom feeds, and X query row sets. to provide unified view.
- Provides data Mashup services capabilities for SOA using the Open ESB and Net Beans infrastructure.
- Exposes the aggregation of multiple data sources to Mashup client frameworks, enabling Web 2.0 type applications in an enterprise.
- Provides time-bound view caching for improved response times and can be called as a virtual materialized view.
- Provides multiple views of the resulting data through XSLT transformation by applying a composite weaving pattern.
- Transforms the resulting data set into various formats by weaving the output with an XSLT Service Engine, enabling deployment on multiple channels.
- Ensures extensibility through the ability to consume JBI Services.

If using SUN Data Mashup Engine design, the time component allow me to specify the data sources and target database for the system. Design-time components include the Net Beans project system, a wizard to guide me to create and configure Data Mashup process, and a mapping editor where i can easily can map source and target data and customize the transformation. The output of a Data Mashup process can be further weaved with the XSLT Service Engine to produce different output formats that can be deployed to multiple channels. The following diagram (figure 3.3) shows the Sun Data Integrator components and their relationship to one another. [16]

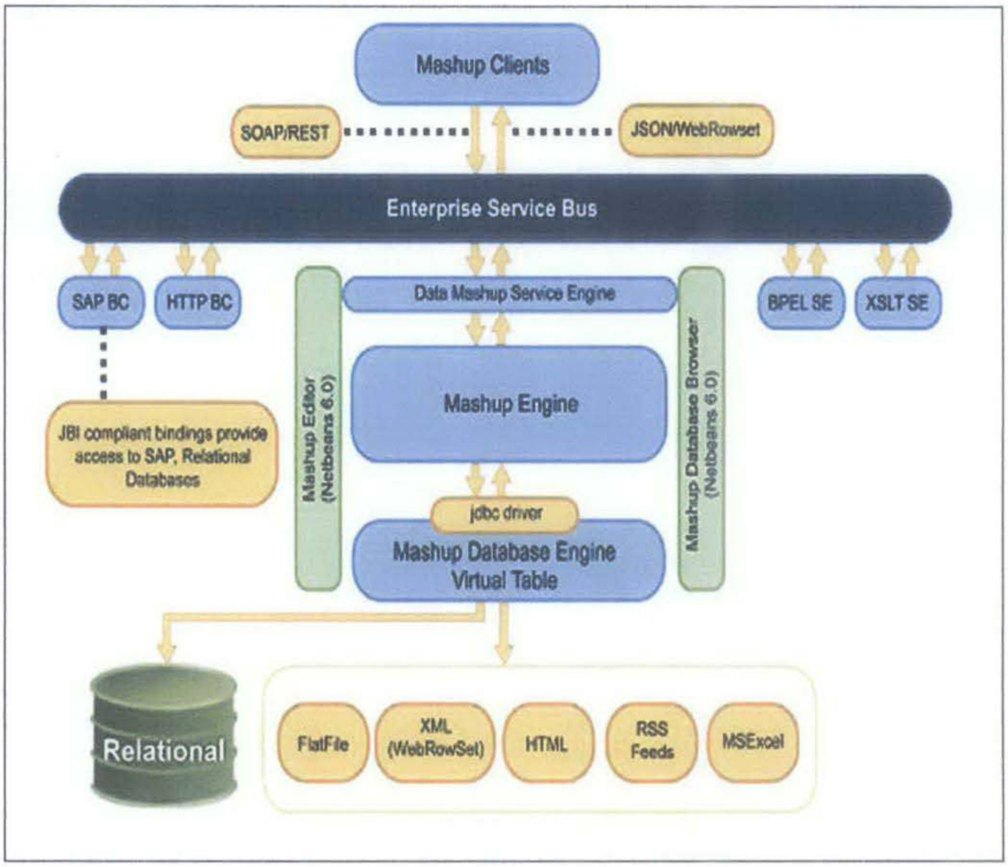

Mashup Clients
SOAP/REST
JSON/WebRowset
Enterprise Service Bus
SAP BC
HTTP BC
Data Mashup Service Engine
BPEL SE
XSLT SE
Mashup Editor (Netbeans 6.0)
Mashup Engine
Mashup Database Browser (Netbeans 6.0)
JBI compliant bindings provide access to SAP, Relational Databases
jdbc driver
Mashup Database Engine Virtual Table
Relational
FlatFile
XML (WebRowSet)
HTML
RSS Feeds
MSExcel

Figure 2.1 :SUN Data Mashup Architecture [16].

2.4 SAMPLE SYSTEM LAYOUT SIMILAR TO MIP

2.4.1 Wikimapia

WikiMapia is a privately owned, online map and satellite imaging resource that combines Google Maps with a wiki system, allowing users to add information, in the form of a note, to any location on Earth. The Goal of Wikimapia is to create and maintain a free, complete, multilingual, up-to-date map of the whole world [17].

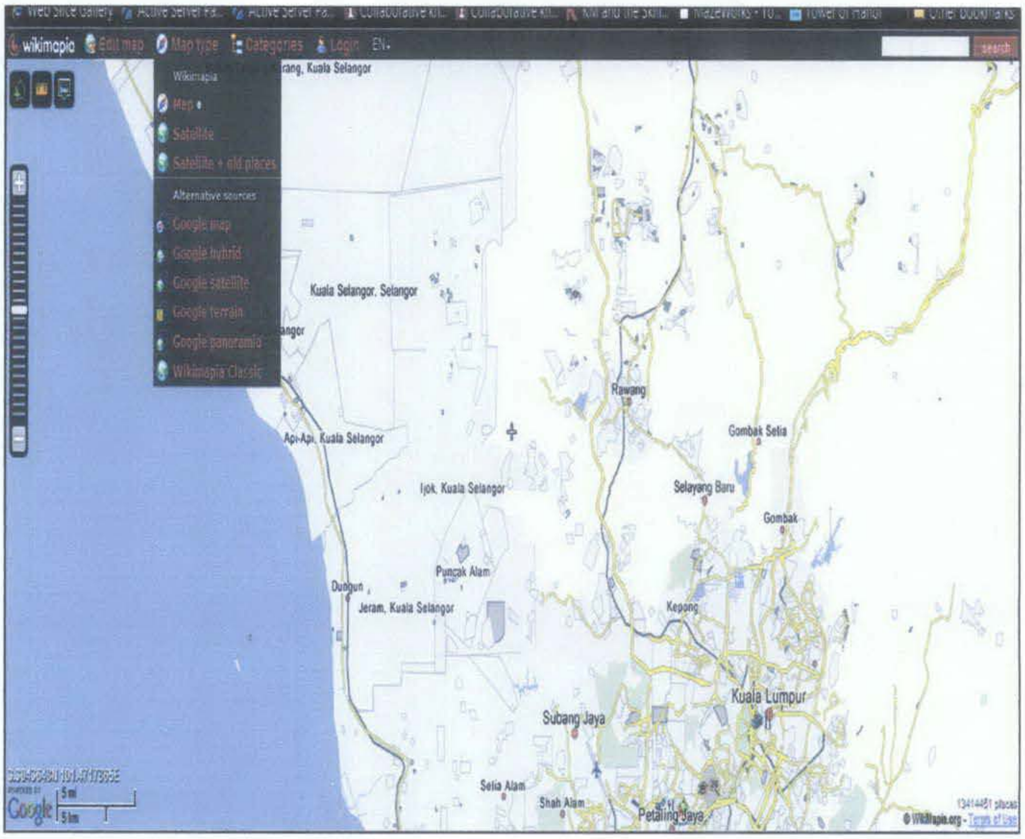


Figure 2.2: Wikimapia layout application [17]

2.4.2 HealthMap

HealthMap brings together disparate data sources to achieve a unified and comprehensive view of the current global state of infectious diseases and their effect on human and animal health. This freely available Web site integrates outbreak data of varying reliability, ranging from news sources (such as Google News) to crated personal accounts (such as ProMED) to validated official alerts (such as World Health Organization). Through an automated text processing system, the data is aggregated by disease and displayed by location for user-friendly access to the original alert. HealthMap provides a jumping-off point for real-time information on emerging infectious diseases and has particular interest for public health officials and international travellers [18].

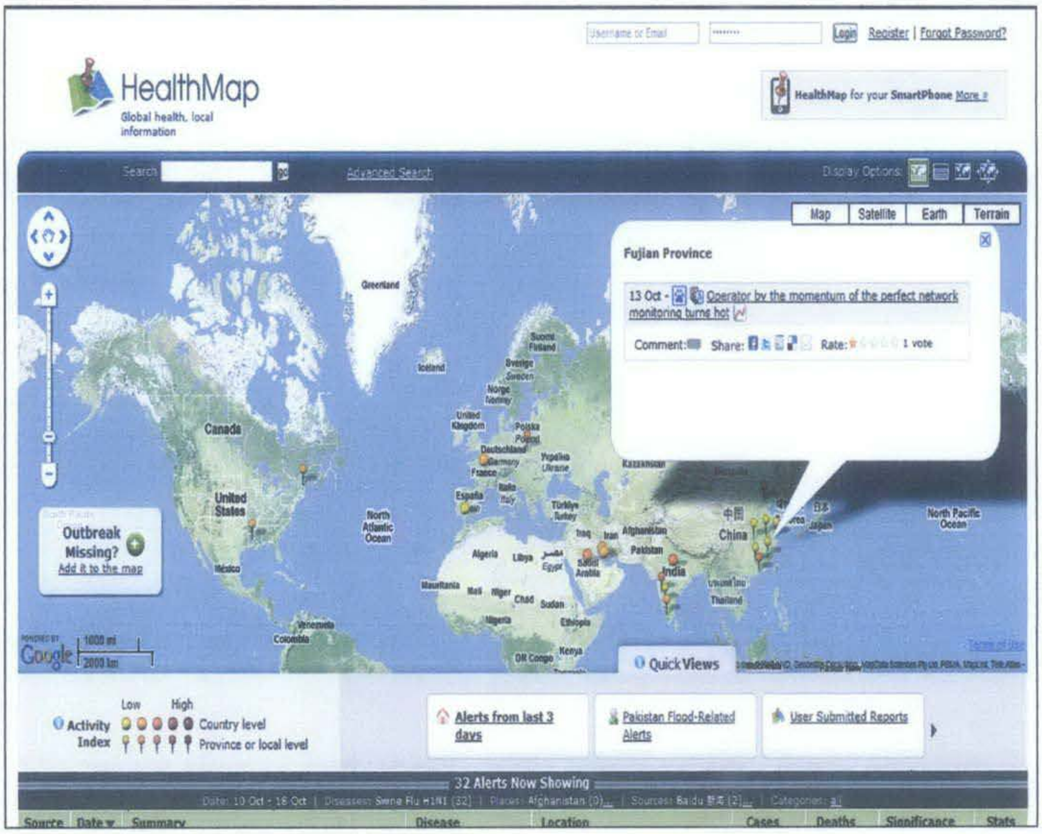


Figure 2.3: Health Map website layout [18].

2.4.3 Geocommons – Google Maps Meet Heat Maps

The maps themselves are based on Google maps, but the data is presented in the form of shapes and heat maps, which is visually very effective, but unfortunately, in practice it's sometimes a bit slow.

GeoCommons is quite an advanced service, and although it can be used for fun, businesses, professionals, and geoenthusiasts will make more use of it than casual users. It takes a while to get used to all that it can offer, but once you do, the results can be very rewarding [21].

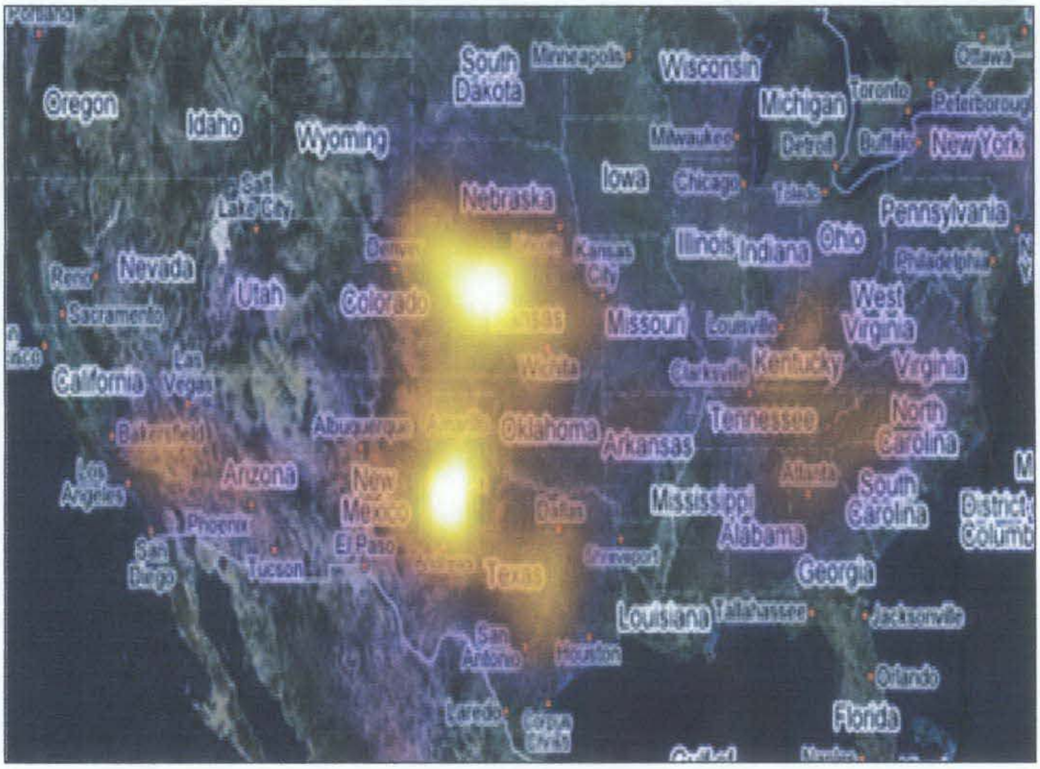
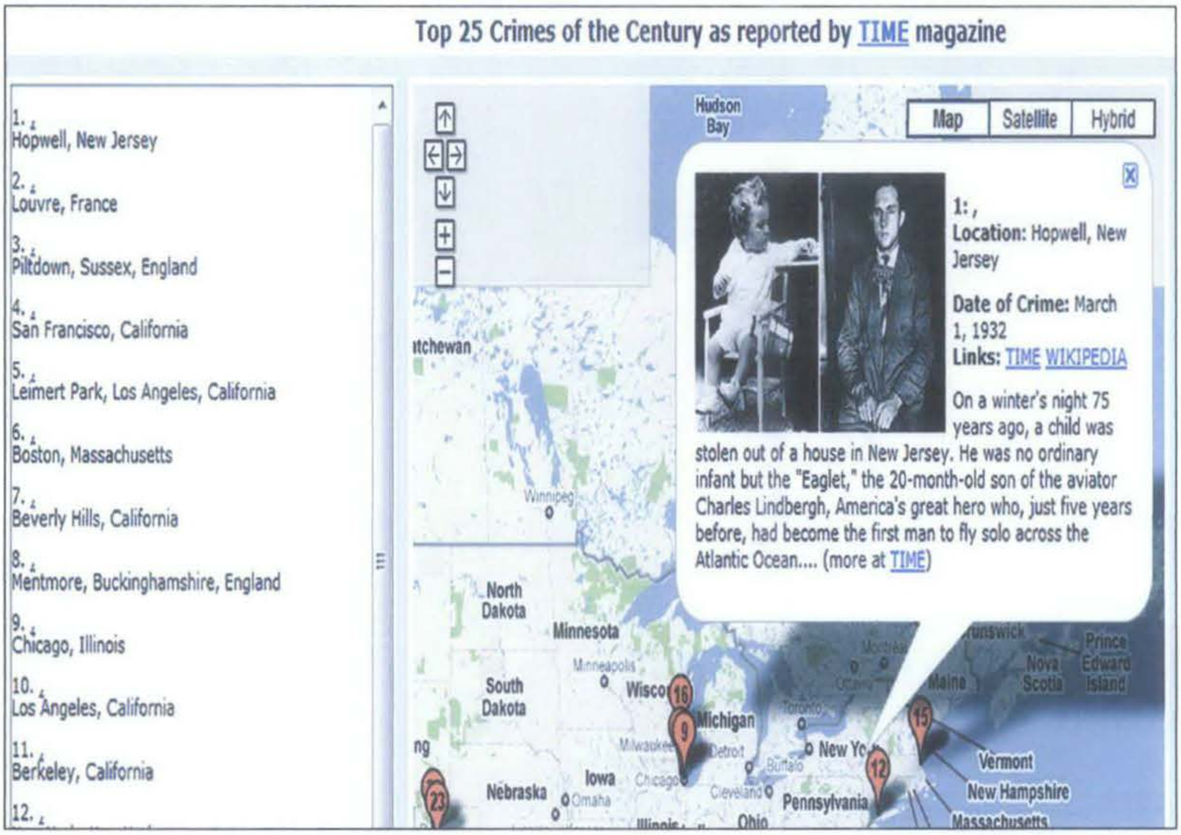


Figure 2.4 : Geocommons of heat maps [21]

2.4.4 MIBAZAAR

A weblog that contains a collection of Google Maps Mashup. Figure 2.5 shows one of interesting mashup in MIBAZAAR.



Top 25 Crimes of the Century as reported by TIME magazine

1.
Hopwell, New Jersey

2.
Louvre, France

3.
Piltdown, Sussex, England

4.
San Francisco, California

5.
Leimert Park, Los Angeles, California

6.
Boston, Massachusetts

7.
Beverly Hills, California

8.
Mentmore, Buckinghamshire, England

9.
Chicago, Illinois

10.
Los Angeles, California

11.
Berkeley, California

12.

1: ,
Location: Hopwell, New Jersey
Date of Crime: March 1, 1932
Links: TIME WIKIPEDIA

On a winter's night 75 years ago, a child was stolen out of a house in New Jersey. He was no ordinary infant but the "Eaglet," the 20-month-old son of the aviator Charles Lindbergh, America's great hero who, just five years before, had become the first man to fly solo across the Atlantic Ocean.... (more at TIME)

Figure 2.5: Mashup of top 25 crimes of the century in MIBAZAAR [22].

CHAPTER 3

METHODOLOGY TO BE USED

3.1 PROJECT METHODOLOGY

In many ways building this system required four fundamental phases which are planning, analysis, design and implementation. This chapter focuses on the methodology used to build the system application. The Mashup in this case use prototyping-based methodology which falls under Rapid Application Development category.

A prototyping-based methodology performs analysis, design, and implementation phase concurrently, and all three phases are performed repeatedly in a cycle until the system is completed. With these methodologies, the basics of analysis and design are performed, as for the Mashup [7], I have done a lot of analysis based on other research related to Mashup and analysis on why the system is needed to help current student to make decision to choose the company for their industrial training. Using this methodology work immediately begins on a system prototype; a 'quick-and-dirty' program provides a minimal amount of features.

The first prototype is usually the first part of the system that the user will use. This is shown the users and project sponsor who provide comments which are used to re-analyze, re-design, and re-implement a second prototype that provides few more features. This process continues until the prototype provides enough functionality to be released and used. After the system is released, refinement occurs until it is accepted as the new system (see Figure 3.1)

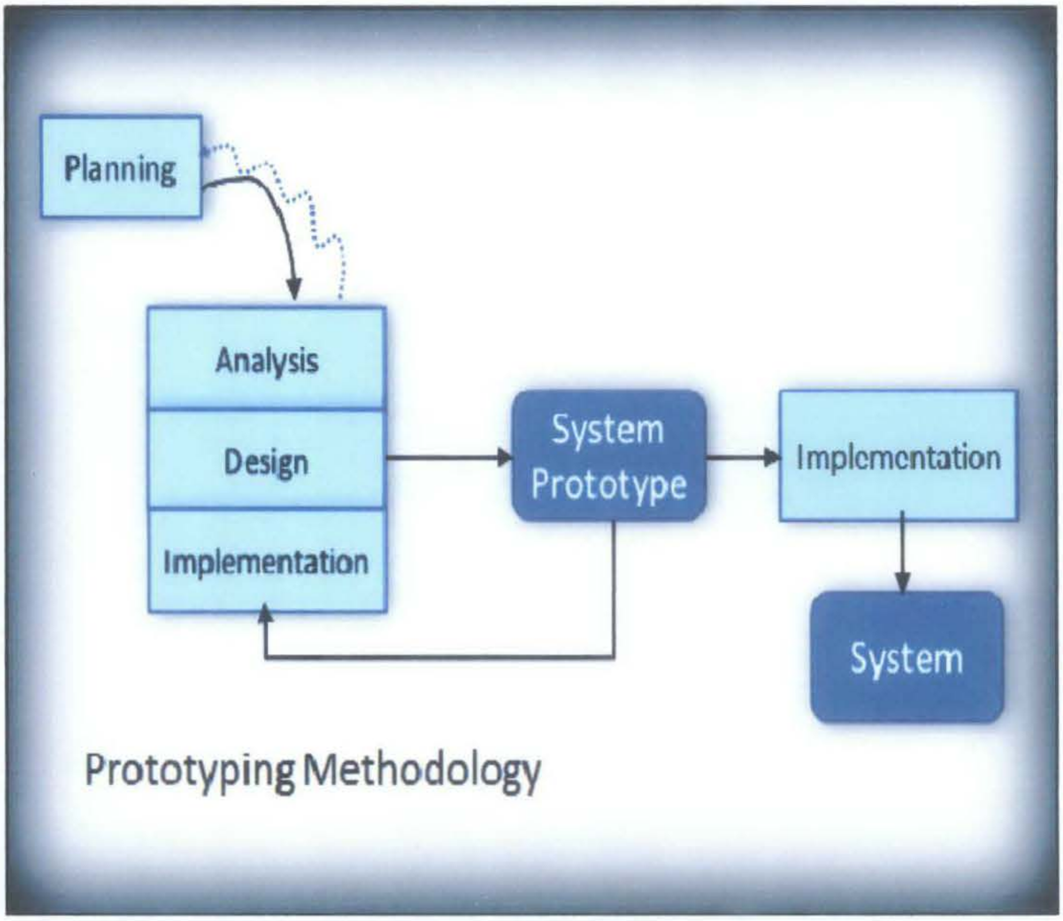


Figure 3.1: The prototyping-based methodology which falls under RAD category.

3.1.1 Planning

In the planning phase, a work plan and Gantt chart (see appendix 1) have been done to put a direct to the project in system development life cycle. A pre-survey also has been distributed to students and staff to identify the system value to the organization. After getting the result of pre-survey than proceed to the next of system development life cycle which is analysis phase.

3.1.2 Analysis

The analysis phase answers the question of who will use the system, what the system will do and where and when it will be used. During this stage, I have investigate the current system and come out with the problem highlights and then ways to design a new system. After all requirements is gathered through interviews and questionnaires. The analysis of this information will leads to the system concept that describes how the system will operate. Then the analyses, system concept and models are combined into document. The result of pre-survey proved that users need this Mashup application which most of them agreed that the Mashup of Internship Placement will help them to make a better decision. The pre-survey is distributed to 30 people and as the result (See Figure 3.2).

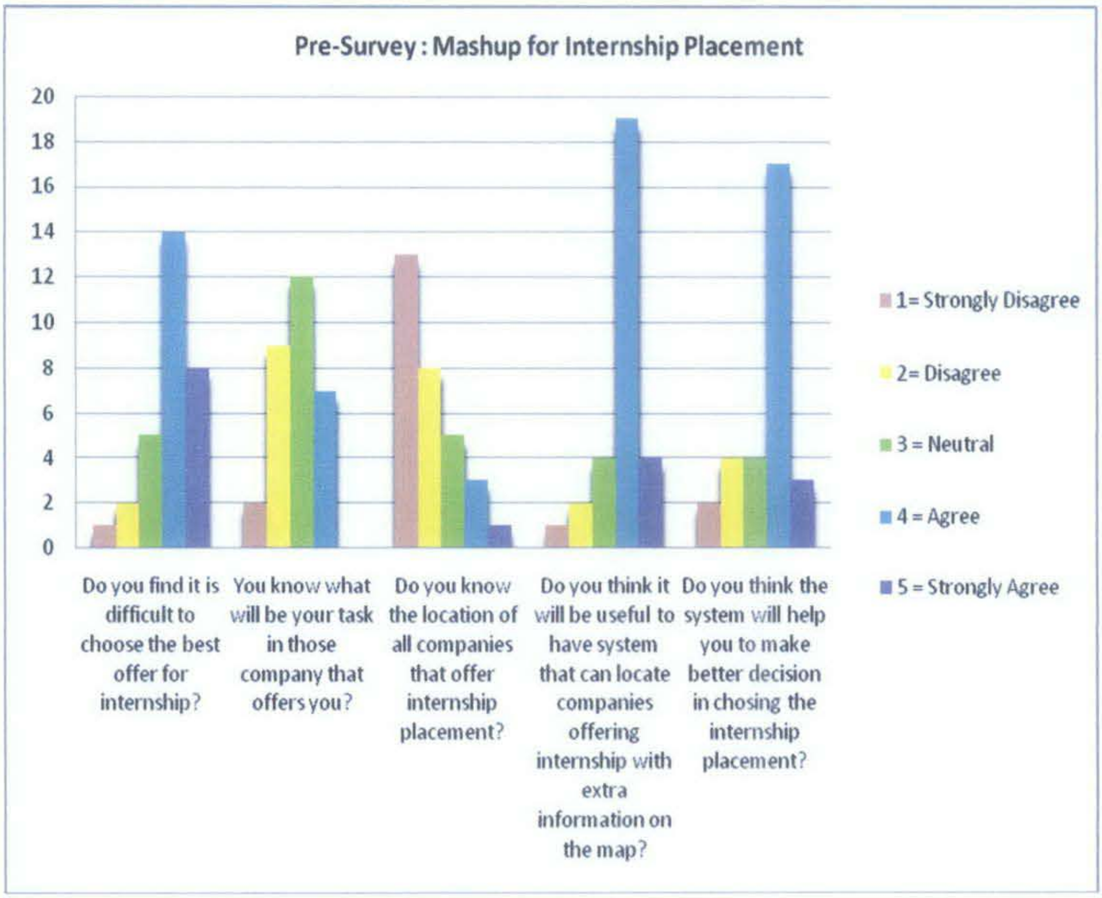


Figure 3.2: Result from Pre-survey of MIP

3.1.3 Design : Proposed Mashup Architecture

The design phase decide how the system will operate in terms of the hardware, software, and network infrastructure; the user interface and the specific databases and files that will be needed. In this stage, the basic architecture design for the system that describes the hardware, software and network infrastructure that will be used is developed [23]. In this case, Mashup is developed using Yahoo Pipes widget. Figure 3.3 show the proposed architecture of typical Mashup application.

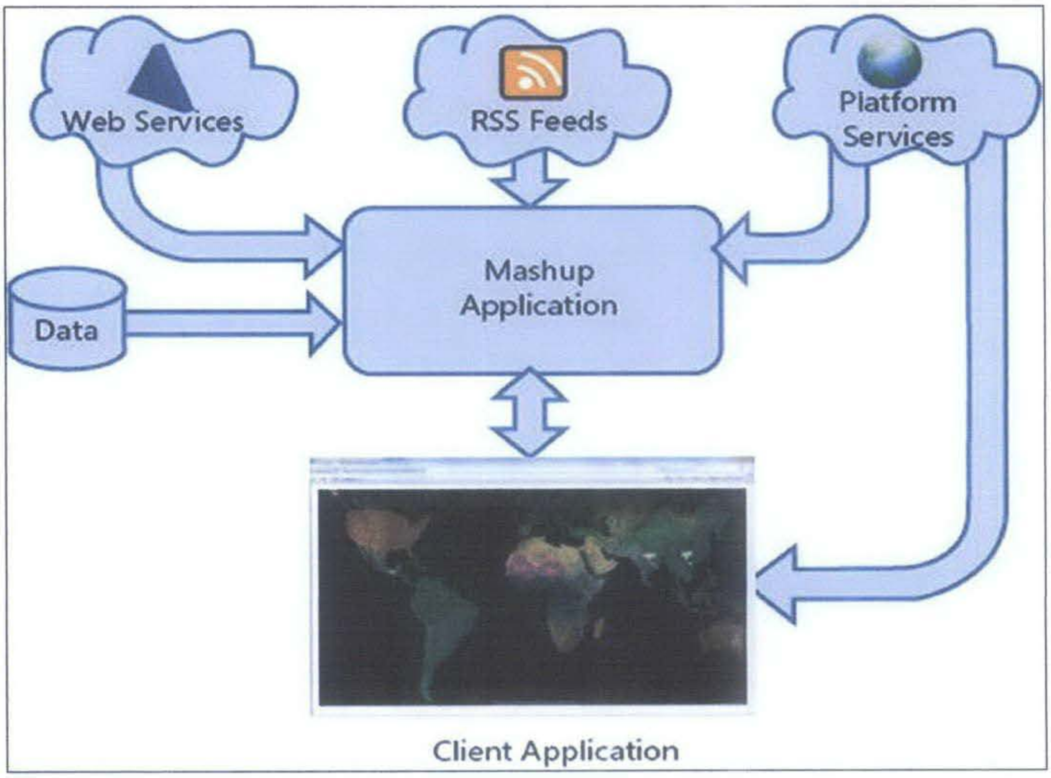


Figure 3.3: The proposed architecture of typical Mashup application [23].

Data

The core element of any Mashup is the data being aggregated and presented to the user [23]. In this paper, data represented the data from database created using Google Maps application. Yahoo Pipes does not require a database that is local to it, the data can strictly come from web service in this case is Google Maps services where data is serialized to XML or JSON (the most common pattern in Internet-based Mashups).

RSS feeds

Use of RSS (Really Simple Syndication) feeds is a common source of primary or supplemental data for Mashups. MIP (Mashup for Internship Placement) use RSS feeds from Google Maps to be extracted by Yahoo Pipes widget. RSS feeds are easy to consume as they are XML documents and many libraries exist to manipulate the feeds [23].

Web Services

It is also common to include calls to Web services within Mashups. It is common to see both WSDL-based Web services and REST-based Web services, with some exposing both styles. Web services can be used to provide additional data or used to transform the data being mashed up. For example in MFIP, data being mashed up using Google Maps and multiple website of companies listed for internship placement. The data for a map-based Mashup may only contain street addresses and a call to a WSDL or REST-based Web services may be required to translate the street address to a Longitude/Latitude coordinate for the map [23].

Platform services

There is special class of services that are used to create Mashups in this case is Yahoo Pipes. Yahoo Pipes is called a platform service because it provides functionality beyond the typical request/response model of traditional web services. In this paper, the mapping service provided by Google Maps and Yahoo Maps. Yahoo Pipes is the platform that has widget to integrate data from Google Maps with Yahoo Maps.

Mashup applications

As we have identified many types of services that can be used to create Mashups, but not yet addressed the importance of the software that creates and delivers the Mashup experience. Yahoo Pipes has that importance in creating the Mashup application. As a result user can use the Mashup application to view the location of companies with extra information on the map.

Client application

The client application is how the Mashup is delivered and presented to the user, For public Internet Mashups the most common client application is a Web browser that receives HTML and Java Script delivered from a Web Server over HTTP. In this case Mashup application from Yahoo Pipes can be embedded into client application using RSS feeds, PHP, JSON, and also as a badge for Yahoo and Google. (See Figure 3.4)

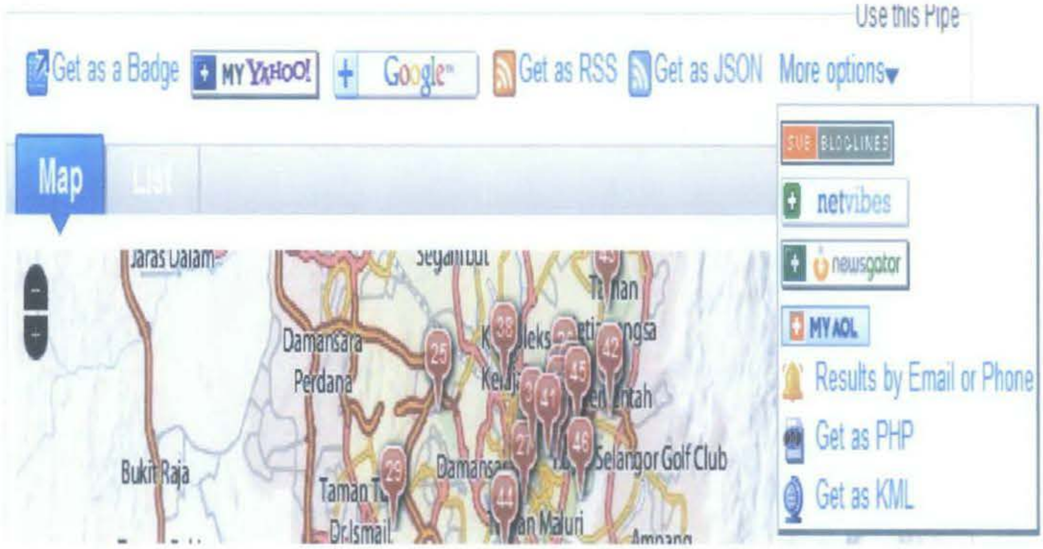
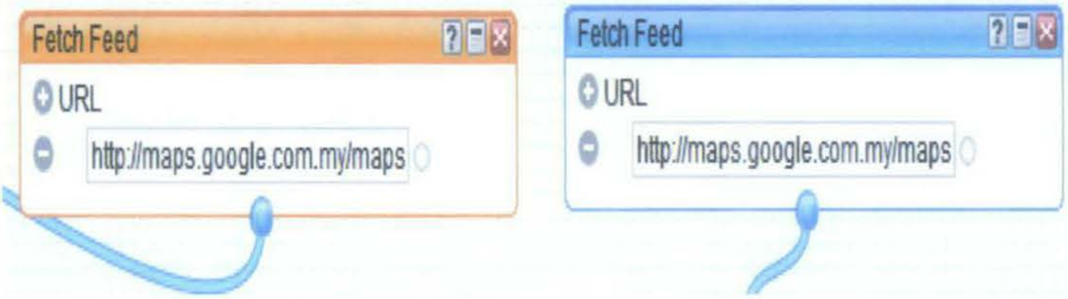


Figure 3.4: Mashup application from Yahoo Pipes that can be embedded into client application

3.1.4 Implementation

The final phase in the SDLC is the implementation phase, during which the system is actually built. In this project, Mashup was build using Yahoo Pipes widget by integrating the pipes with data from multiple sources. The database is created using Google Maps application that allow developer to marked the location of company and link it with companies website. The integration of data in database is using RSS feeds and it is live update. So if the database change, data in Yahoo Pipes will automatically updated. Yahoo Pipes will extract the RSS feeds from Google Maps using fetch feed widget. (See Figure 3.5).



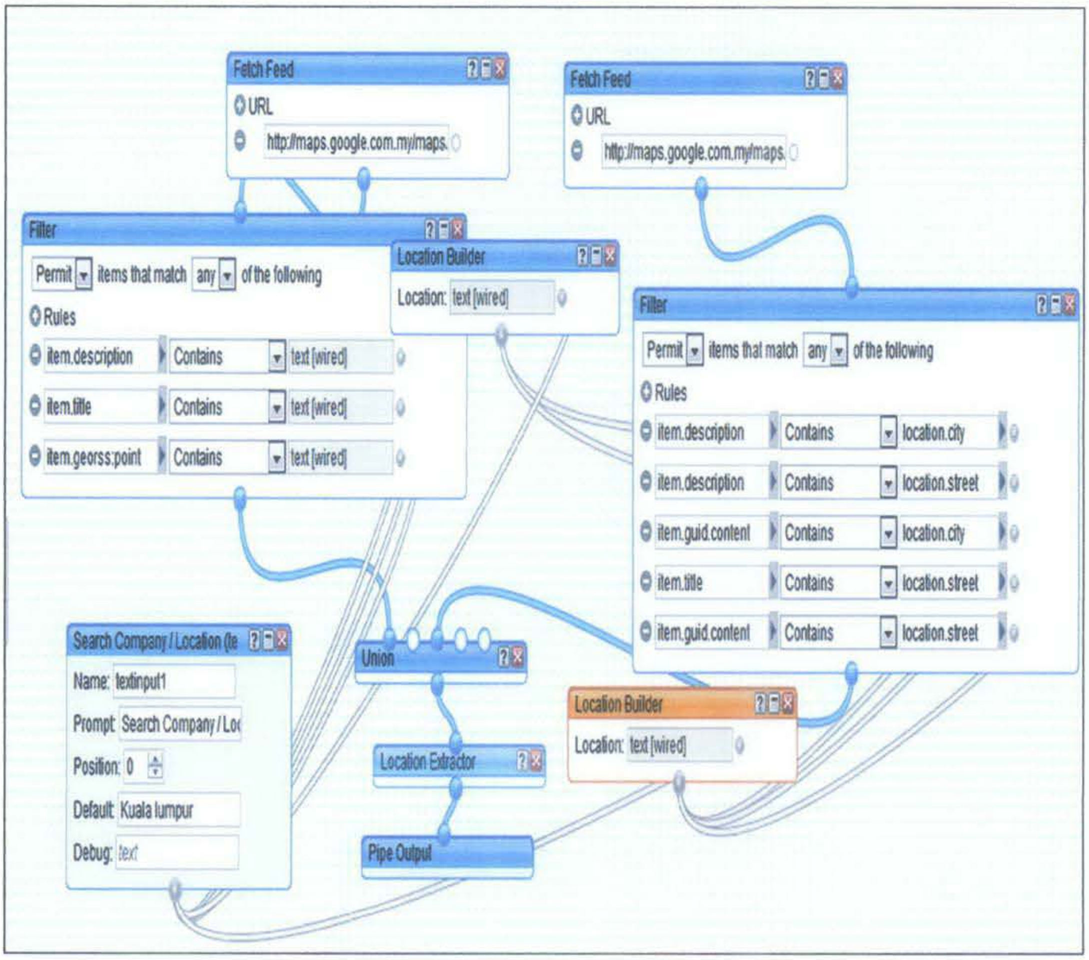
URL: http://maps.google.com.my/maps

URL: http://maps.google.com.my/maps

Figure 3.5: Fetch Feed widget from Yahoo Pipes which extract the RSS feeds

3.2 PROPOSED MASHUP ENGINE USING YAHOO PIPES

Creating Mashup for Internship Placement used several modules or widget in Yahoo Pipes such as fetch data feed module which falls under source, location module which is the location builder widget, user input module, operator modules: union module, filter module and location extractor module (See Figure 3.6).



URL http://maps.google.com.my/maps.

URL http://maps.google.com.my/maps.

Permit items that match any of the following

Rules

item.description Contains text [wired]

item.title Contains text [wired]

item.georss:point Contains text [wired]

Location: text [wired]

Permit items that match any of the following

Rules

item.description Contains location.city

item.description Contains location.street

item.guid.content Contains location.city

item.title Contains location.street

item.guid.content Contains location.street

Name: textinput1

Prompt: Search Company / Lo(

Position: 0

Default: Kuala lumpur

Debug: text

Location: text [wired]

FIGURE 3.6: Proposed Mashup engine using Yahoo Pipes widget

3.2.1 Source

Most Pipes begin with a data source. These modules grab data from somewhere on the internet and bring it into the Pipe for processing. One of the source modules to create Mashup for internship is Fetch Feed module [24].

Fetch Feed module

The Fetch Feed module lets me specify one or more RSS news feeds as input to the Pipe. The module understands feeds in RSS, Atom, and RDF formats. Feeds contain one or more items, when add more feed URLs it will get a single feed combining all the items from the individual feeds.URLs can also be "wired in" to the module from other modules (like the URL Builder) [24]. Fetch feed module extract data from the RSS feed of database from Google Maps.

3.2.2 Operator modules

Filter Module

When there are more data needed to build the Mashup, the filter module allow developer to include or exclude items from a feed. Filter will create rules that compare feed elements to values that is specify for the Mashup. For example, in MIP, I create rules that say " permit items where the description contains the text input from user 'Kuala Lumpur'" or permit items where the item description contains the 'location. city' from user text input. A single Filter module can contain multiple rules and can choose whether an item must match all the rules or if it can just match any rule [24].

Location Extractor Module

This module examines the input feed, looking for information that indicates a geographic location. If it finds geographic data, the module adds a y: location element to the output feed. This element contains several sub-elements, including: lat, lon, quality, country, state, city, street and postal. Some of the sub-element may not be include in the resulting feed and this depends on the amount of the information that can be derived from the input feed. Location Extractor gleans location data from URLs in this Mashup is from maps.google.com. If a y: location element is found in the pipe output, the Pipe Preview feature will optionally display the feed on an interactive map [24].

Union Module

This module merges up to five separate sources of items into a single list of items. In the Figure 3.6 shows that union merge the filter from two items into one item so that it can be link to the location extractor.

3.2.3 User Input Module

User input module add flexibility to the pipes. User can choose to put the text input or location input. This input let user to feed input information into a complete Pipe when user run it.

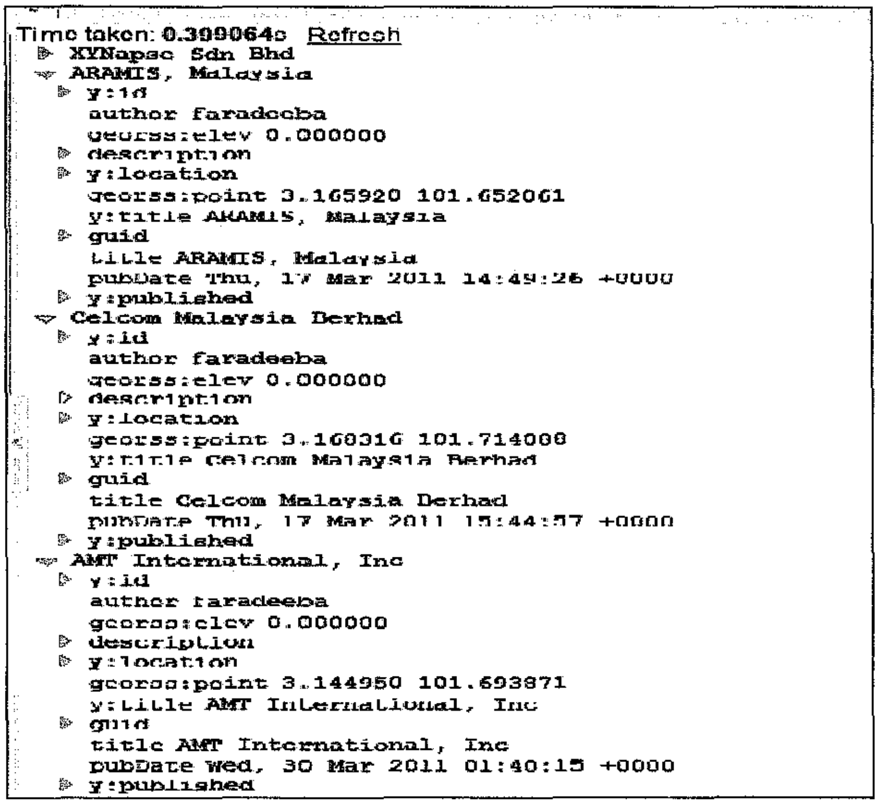
Text Input Module

For MIP I used text input module to let user input the company name or location of the company. Many of other module can accept text, including all the String modules, almost all of the source modules (except Fetch Feed), Filter, For Each: Annotate, Regex, and Rename [24].

3.2.4 Location Module

Location Builder Module

This module converts a description of a place into geographical data. It recognizes addresses, zip codes, airport codes, city/country names, and U.S city/state names. The module outputs a location structure separate the fields for city, state, country, latitude, and longitude. This location can be wired into any module that accepts location types [24].



```
Time taken: 0.399064s  Refresh
 ▷ XYNapse Sdn Bhd
 ▽ ARAMIS, Malaysia
   ▷ y:id
     author faradeeba
     georss:elev 0.000000
   ▷ description
   ▷ y:location
     georss:point 3.165920 101.652061
     y:title ARAMIS, Malaysia
   ▷ guid
     title ARAMIS, Malaysia
     pubDate Thu, 17 Mar 2011 14:49:26 +0000
   ▷ y:published
 ▽ Celcom Malaysia Berhad
   ▷ y:id
     author faradeeba
     georss:elev 0.000000
   ▷ description
   ▷ y:location
     georss:point 3.160316 101.714008
     y:title Celcom Malaysia Berhad
   ▷ guid
     title Celcom Malaysia Berhad
     pubDate Thu, 17 Mar 2011 15:44:57 +0000
   ▷ y:published
 ▽ AMT International, Inc
   ▷ y:id
     author faradeeba
     georss:elev 0.000000
   ▷ description
   ▷ y:location
     georss:point 3.144950 101.693871
     y:title AMT International, Inc
   ▷ guid
     title AMT International, Inc
     pubDate Wed, 30 Mar 2011 01:40:15 +0000
   ▷ y:published
```

Figure 3.7: The debugger pipe output.

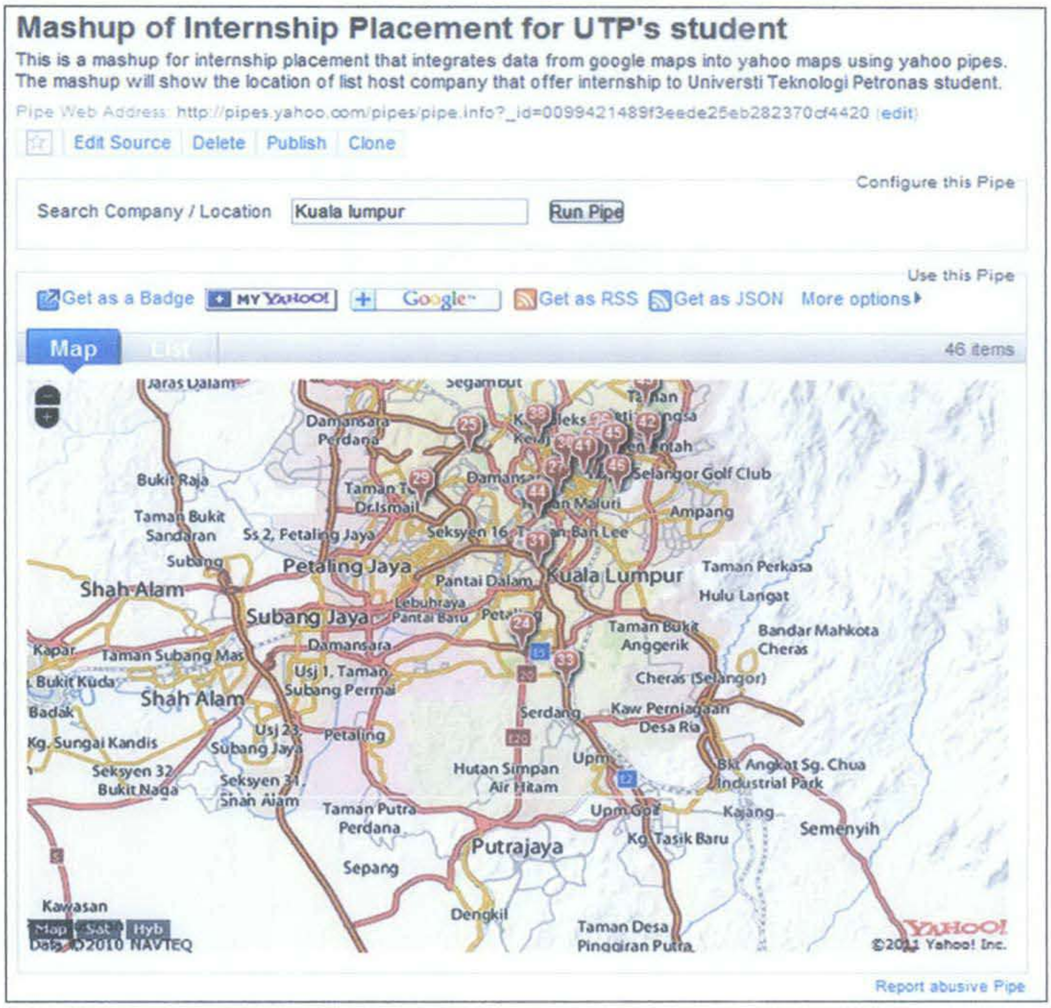
Using all linked Pipes Mashup engine can be run and the output will show the debugger below the widget (see Figure 3.7) which capture the list of companies that offer internship placement from Google Maps. If the debugger does not give any result that means the Pipes is wrongly linked or it is not complete to view the exact result of user input

CHAPTER 4

RESULT AND DISCUSSION

4.1 MASHUP FOR INTERSHIP PLACEMENT LAYOUT

Figure 4.1 to Figure 4.4 is the screen shot of MIP in Yahoo Pipes. Figure 4.1 is the result after user input the text on the search area to find company in Kuala Lumpur. When user clicks the run pipe button, pipes will extract and integrate the data and show the result on the maps.



Mashup of Internship Placement for UTP's student

This is a mashup for internship placement that integrates data from google maps into yahoo maps using yahoo pipes. The mashup will show the location of list host company that offer internship to Universti Teknologi Petronas student.

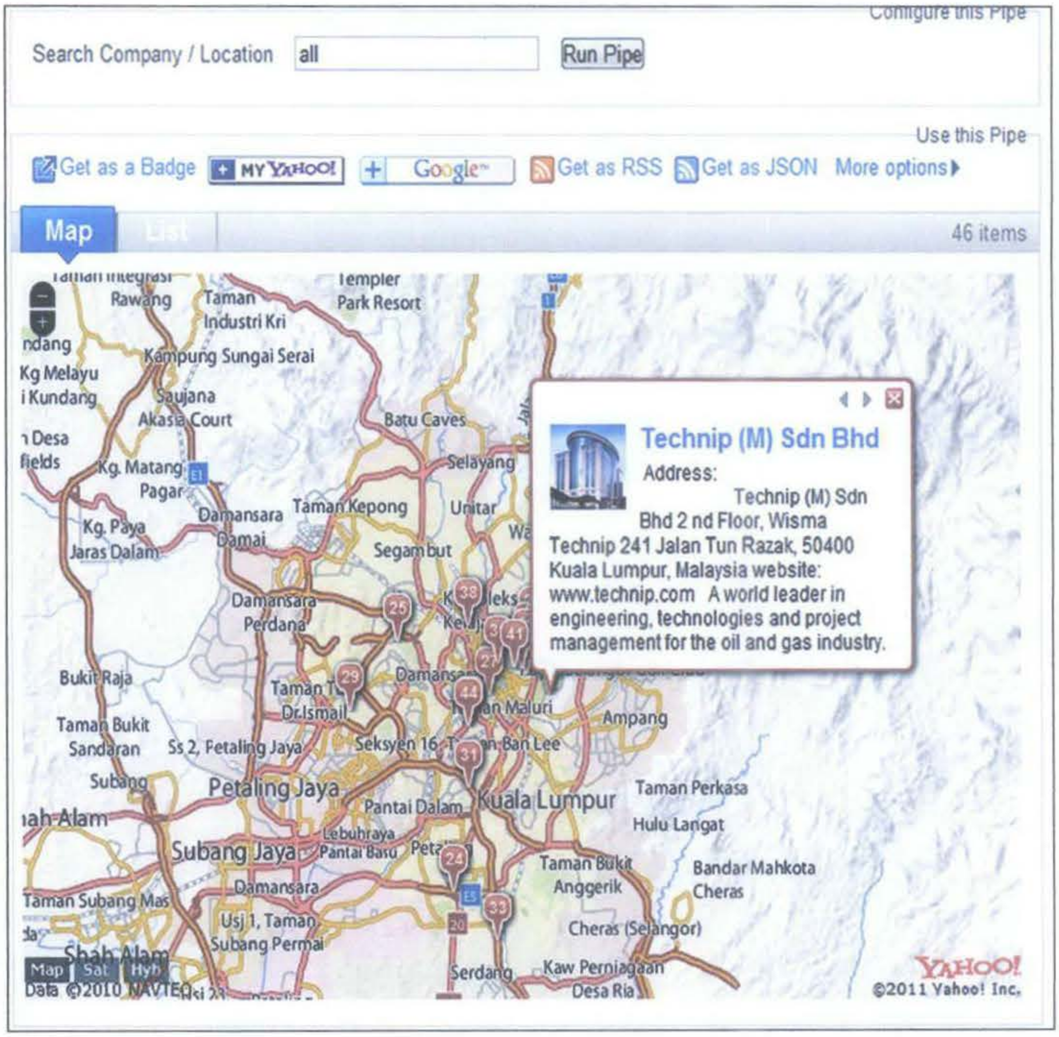
Pipe Web Address: http://pipes.yahoo.com/pipes/pipe.info?_id=0099421489f3eede25eb282370cf4420 (edit)

Search Company / Location: Kuala lumpur

46 items

Figure 4.1: Mashup of companies offered internship in Kuala Lumpur

Figure 4.2 show Mashup of all companies that offer internship. The user search all in the search button. As the result all company in the database will appear on the map. While Figure 4.3 show a company based on user input in the search button. It will show the location of the company and some extra information on the map. Figure 4.4 shows list of companies that offer internship placement using the same Mashup. Yahoo Pipes allow user to view in interactively on the map or to view it in the list format.



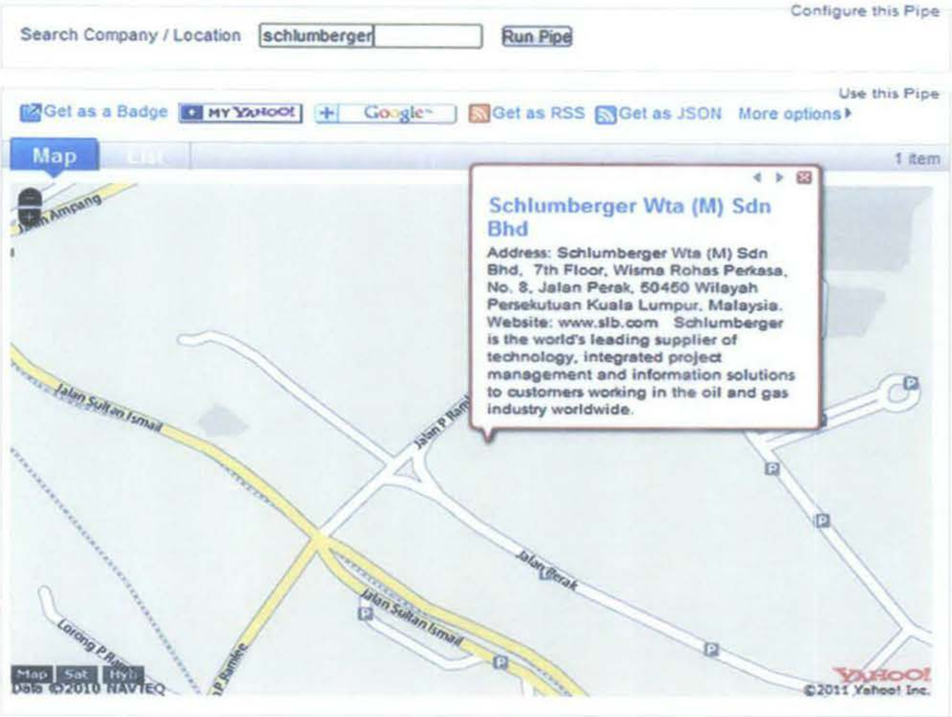
Search Company / Location: all

Technip (M) Sdn Bhd

Address:
Technip (M) Sdn Bhd 2 nd Floor, Wisma Technip 241 Jalan Tun Razak, 50400 Kuala Lumpur, Malaysia website: www.technip.com A world leader in engineering, technologies and project management for the oil and gas industry.

46 items

Figure 4.2: Mashup showing all companies on the map.

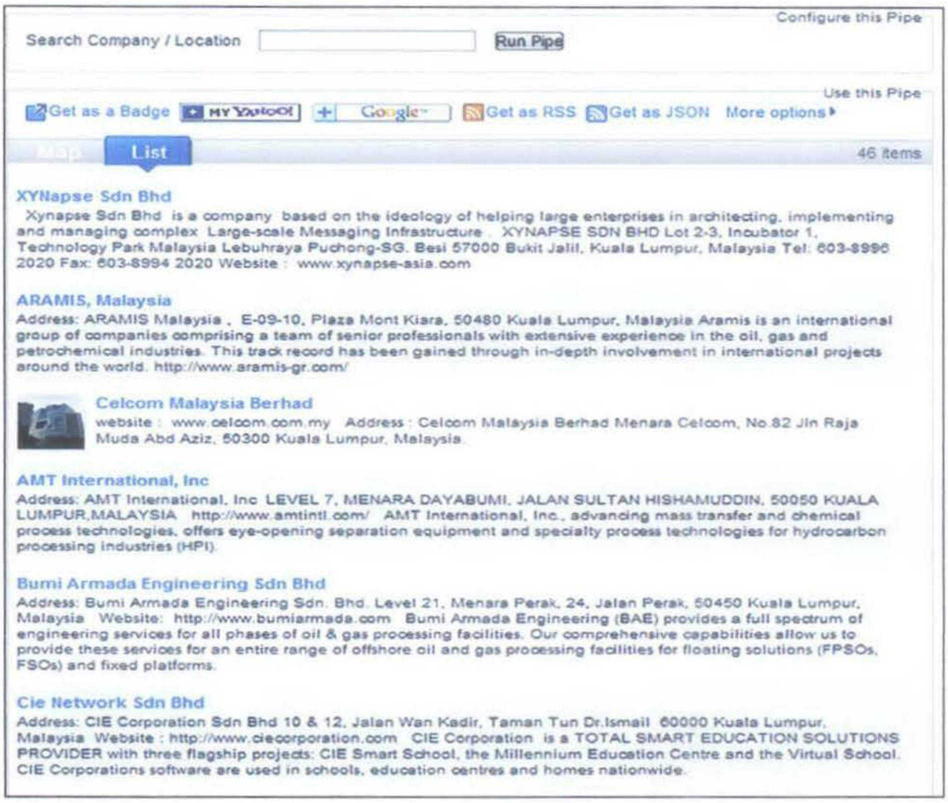


Search Company / Location: schlumberger

Schlumberger Wta (M) Sdn Bhd

Address: Schlumberger Wta (M) Sdn Bhd, 7th Floor, Wisma Rohas Perkasa, No. 8, Jalan Perak, 50450 Wilayah Persekutuan Kuala Lumpur, Malaysia. Website: www.slb.com Schlumberger is the world's leading supplier of technology, integrated project management and information solutions to customers working in the oil and gas industry worldwide.

Figure 4.3: Mashup show location of one company on the map.



Search Company / Location:

XYNapse Sdn Bhd

Xynapse Sdn Bhd is a company based on the ideology of helping large enterprises in architecting, implementing and managing complex Large-scale Messaging Infrastructure . XYNAPSE SDN BHD Lot 2-3, Incubator 1, Technology Park Malaysia Lebuhraya Puchong-SG. Besi 57000 Bukit Jalil, Kuala Lumpur, Malaysia Tel: 603-8996 2020 Fax: 603-8994 2020 Website : www.xynapse-asia.com

ARAMIS, Malaysia

Address: ARAMIS Malaysia , E-09-10, Plaza Mont Kiara, 50480 Kuala Lumpur, Malaysia Aramis is an international group of companies comprising a team of senior professionals with extensive experience in the oil, gas and petrochemical industries. This track record has been gained through in-depth involvement in international projects around the world. http://www.aramis-gr.com/

Celcom Malaysia Berhad

website : www.celcom.com.my Address : Celcom Malaysia Berhad Menara Celcom, No.82 Jln Raja Muda Abd Aziz, 50300 Kuala Lumpur, Malaysia.

AMT International, Inc

Address: AMT International, Inc LEVEL 7, MENARA DAYABUMI, JALAN SULTAN HISHAMUDDIN, 50050 KUALA LUMPUR,MALAYSIA http://www.amtintl.com/ AMT International, Inc., advancing mass transfer and chemical process technologies, offers eye-opening separation equipment and specialty process technologies for hydrocarbon processing industries (HPI).

Bumi Armada Engineering Sdn Bhd

Address: Bumi Armada Engineering Sdn. Bhd. Level 21, Menara Perak, 24, Jalan Perak, 50450 Kuala Lumpur, Malaysia Website: http://www.bumiarmada.com Bumi Armada Engineering (BAE) provides a full spectrum of engineering services for all phases of oil & gas processing facilities. Our comprehensive capabilities allow us to provide these services for an entire range of offshore oil and gas processing facilities for floating solutions (FPSOs, FSOs) and fixed platforms.

Cie Network Sdn Bhd

Address: CIE Corporation Sdn Bhd 10 & 12, Jalan Wan Kadir, Taman Tun Dr.Ismail 60000 Kuala Lumpur, Malaysia Website : http://www.ciecorporation.com CIE Corporation is a TOTAL SMART EDUCATION SOLUTIONS PROVIDER with three flagship projects: CIE Smart School, the Millennium Education Centre and the Virtual School. CIE Corporations software are used in schools, education centres and homes nationwide.

Figure 4.4: Mashup viewed in list format.

4.2 SYSTEM PERFORMANCE EVALUATION

The survey is conducted online using google document. A survey form is created and submitted to 10 users and the outcome of the study prove that this application will help them to make better decision with their placement of internship and also one of interactive application of sharing knowledge. Figure 4.5 show the result of the survey (See apendix 2).

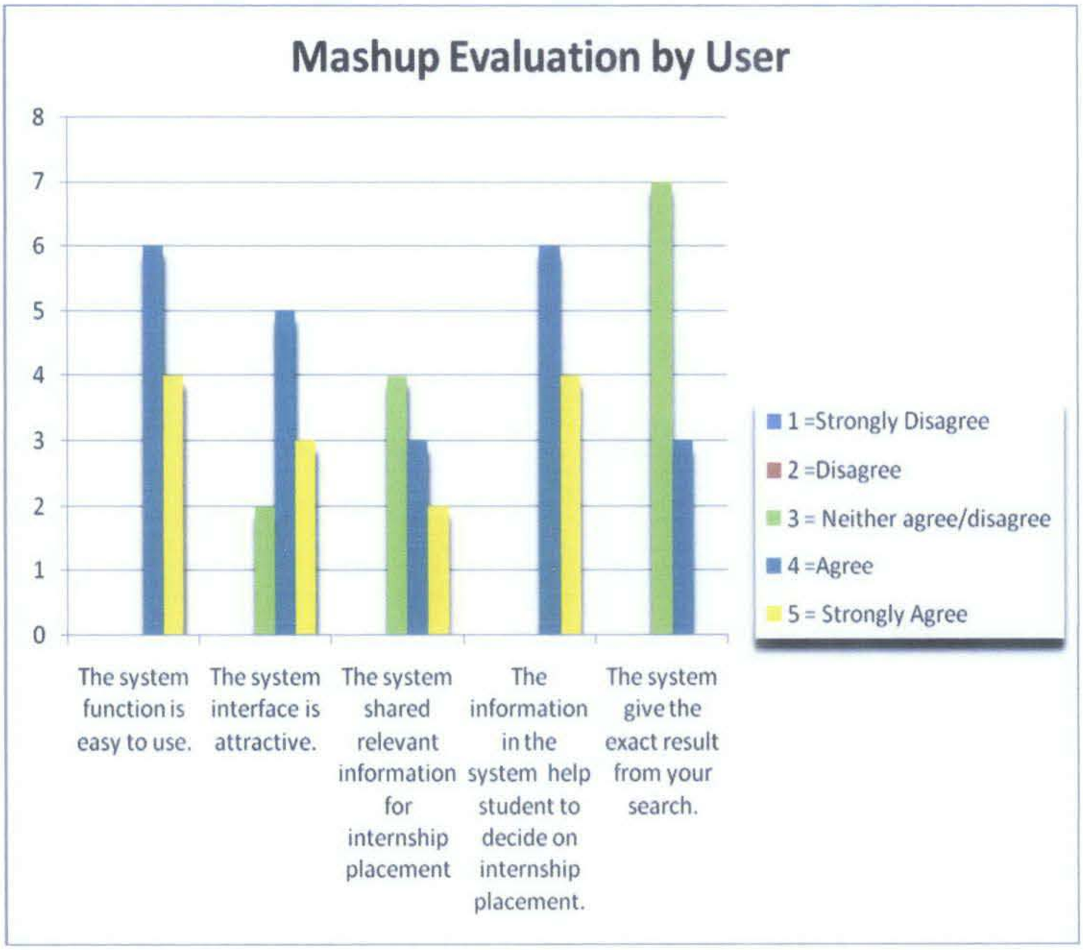
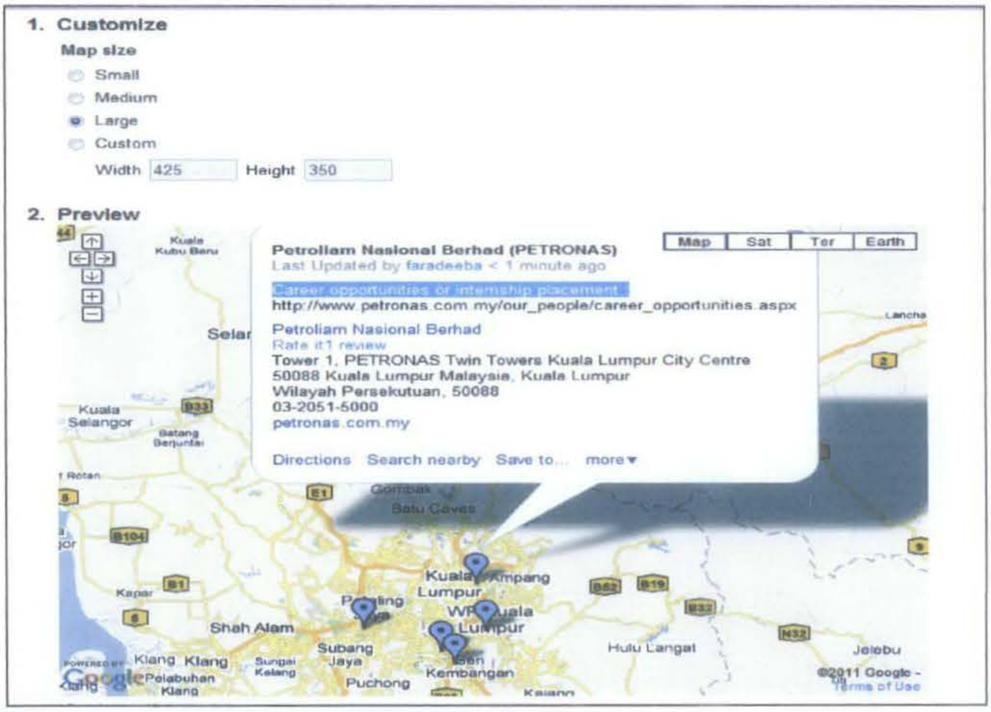


Figure 4.5: System performance evaluation.

4.3 GOOGLE MAP DATABASE

This is the example of internship placement on the map in Selangor and Kuala Kuala Lumpur using Google Maps API. Below is the result (screen shot) of the mashup and the embbeded code that can be use to put into the client application. Figure 4.6 shows how the mashup of the map can be customize into different size and the code will then be embedded into another website.

Figure 4.7 shows the result of embedded code in the internship mashup application. Mashup integrate the location of the company on the maps with the company websites and information about the job or careers opportunites that available in their company. After search the company location a marker will appear on the map and I have integrate it with the url of company's website. When we click at the marker it will show the infomation and the address of the company. A list under Internship placement in Selangor and Kuala Lumpur will appear on the left side of the map.



1. Customize

Map size

○ Small
○ Medium
● Large
○ Custom

Width 425 Height 350

2. Preview

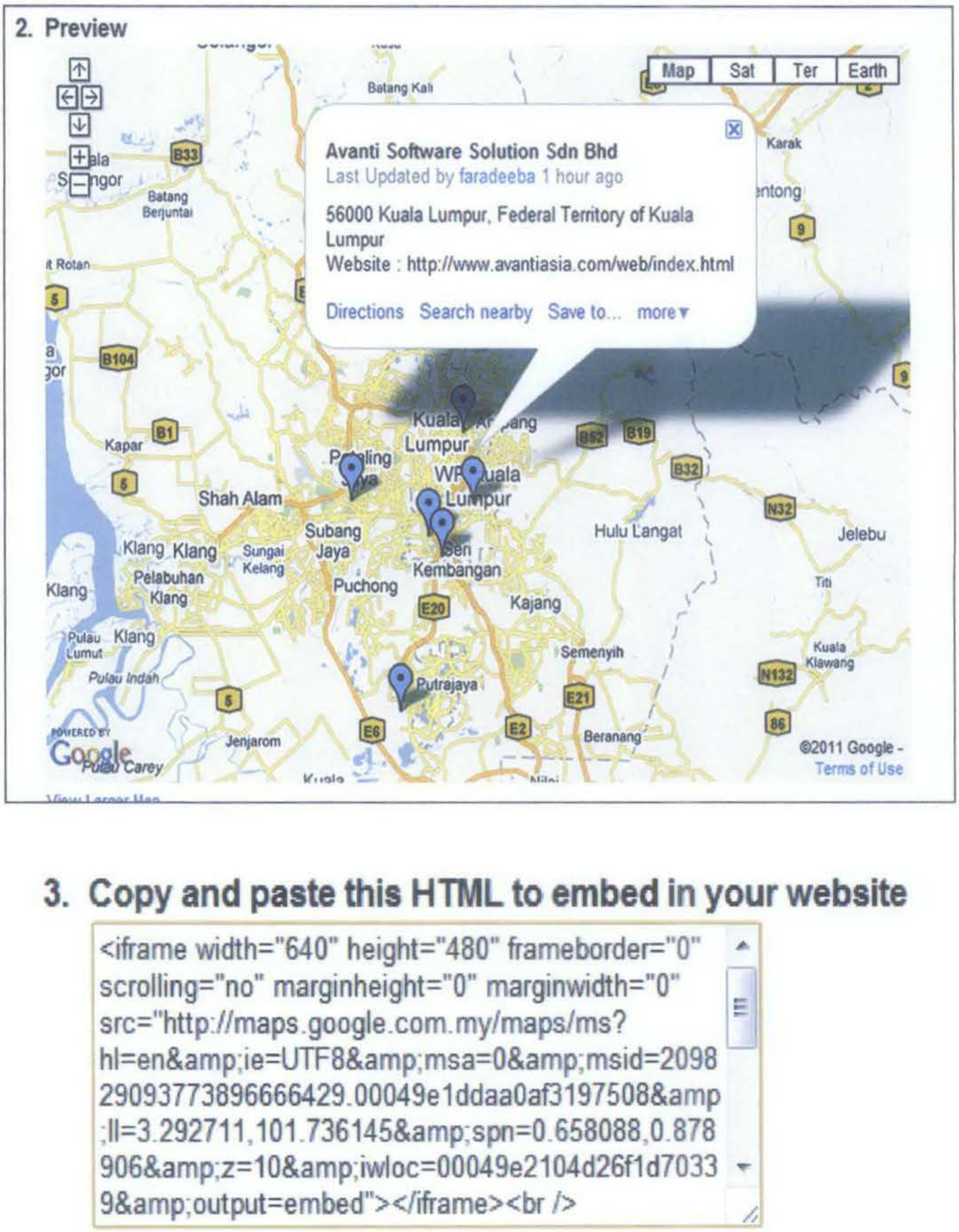
Petroliam Nasional Berhad (PETRONAS)
Last Updated by faradeeba < 1 minute ago

Career opportunities or internship placement
http://www.petronas.com.my/our_people/career_opportunities.aspx

Petroliam Nasional Berhad
Rate it1 review
Tower 1, PETRONAS Twin Towers Kuala Lumpur City Centre
50088 Kuala Lumpur Malaysia, Kuala Lumpur
Wilayah Persekutuan, 50088
03-2051-5000
petronas.com.my

Directions Search nearby Save to... more▾

Figure 4.6 : Database created using Google Map that can be customize [19].



2. Preview

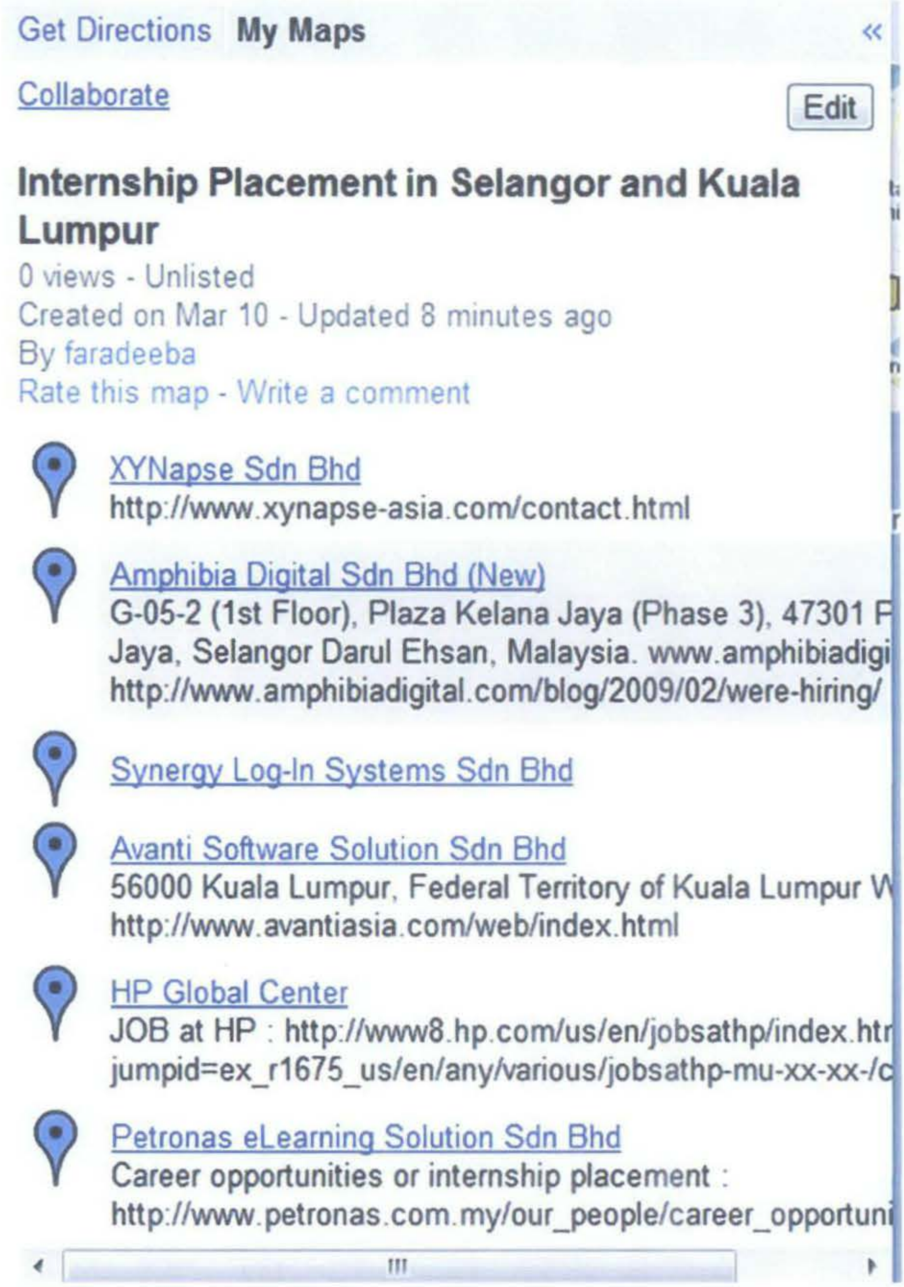
Avanti Software Solution Sdn Bhd
Last Updated by faradeeba 1 hour ago

56000 Kuala Lumpur, Federal Territory of Kuala Lumpur
Website : http://www.avantiasia.com/web/index.html

3. Copy and paste this HTML to embed in your website

```
<iframe width="640" height="480" frameborder="0"
scrolling="no" marginheight="0" marginwidth="0"
src="http://maps.google.com.my/maps/ms?
hl=en&amp;ie=UTF8&amp;msa=0&amp;msid=2098
29093773896666429.00049e1ddaa0af3197508&amp
;ll=3.292711,101.736145&amp;spn=0.658088,0.878
906&amp;z=10&amp;iwloc=00049e2104d26f1d7033
9&amp;output=embed"></iframe><br />
```

Figure 4.7: The HTML code after customize



Internship Placement in Selangor and Kuala Lumpur

0 views - Unlisted
Created on Mar 10 - Updated 8 minutes ago
By faradeeba
Rate this map - Write a comment

XYNapse Sdn Bhd
http://www.xynapse-asia.com/contact.html

Amphibia Digital Sdn Bhd (New)
G-05-2 (1st Floor), Plaza Kelana Jaya (Phase 3), 47301 P
Jaya, Selangor Darul Ehsan, Malaysia. www.amphibiadigi
http://www.amphibiadigital.com/blog/2009/02/were-hiring/

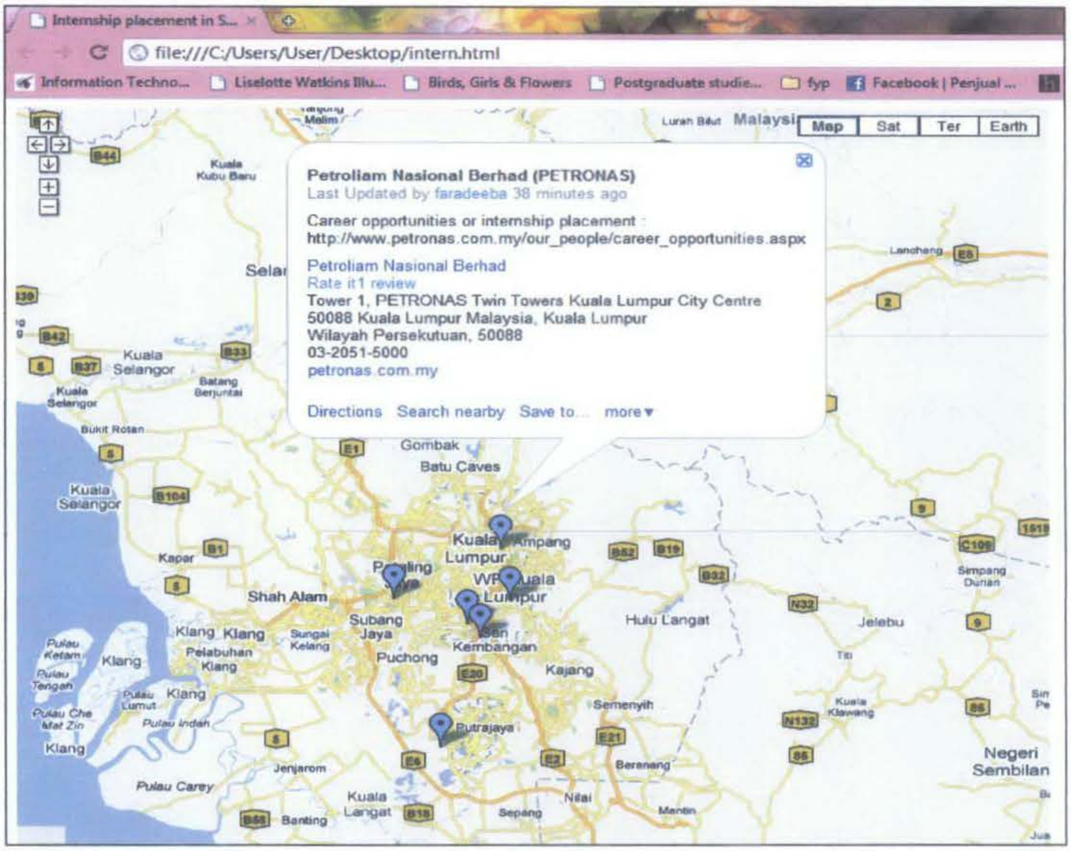
Synergy Log-In Systems Sdn Bhd

Avanti Software Solution Sdn Bhd
56000 Kuala Lumpur, Federal Territory of Kuala Lumpur W
http://www.avantiasia.com/web/index.html

HP Global Center
JOB at HP : http://www8.hp.com/us/en/jobsathp/index.htr
jumpid=ex_r1675_us/en/any/various/jobsathp-mu-xx-xx-/c

Petronas eLearning Solution Sdn Bhd
Career opportunities or internship placement :
http://www.petronas.com.my/our_people/career_opportuni

Figure 4.8: List of Company that offer internship placement for the moment [19].



Petroliam Nasional Berhad (PETRONAS)
Last Updated by faradeeba 38 minutes ago

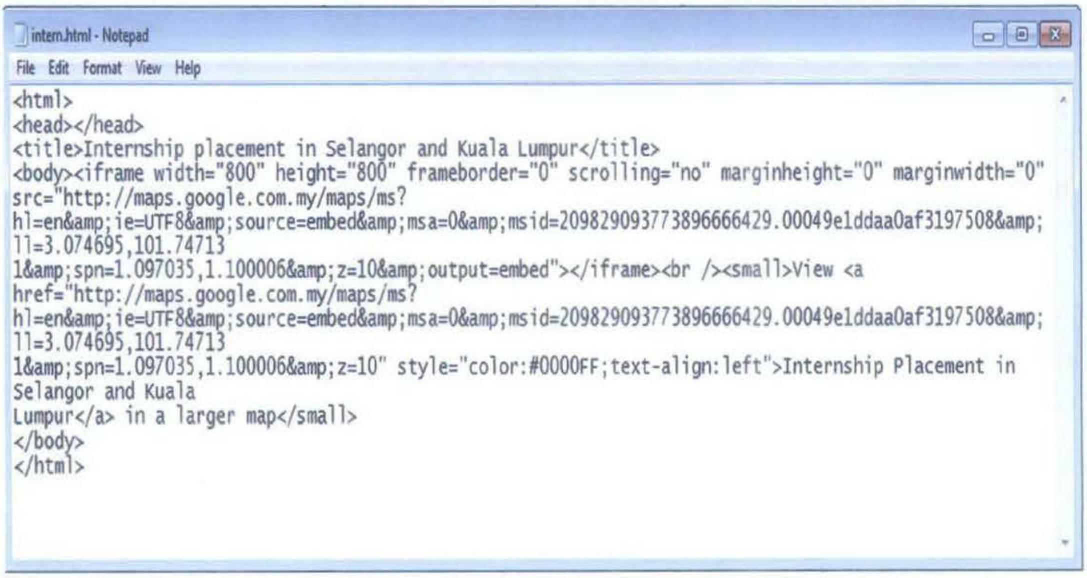
Career opportunities or internship placement :
http://www.petronas.com.my/our_people/career_opportunities.aspx

Petroliam Nasional Berhad
Rate it1 review
Tower 1, PETRONAS Twin Towers Kuala Lumpur City Centre
50088 Kuala Lumpur Malaysia, Kuala Lumpur
Wilayah Persekutuan, 50088
03-2051-5000
petronas.com.my

Directions Search nearby Save to... more ▾

Figure 4.9: Result of Google map Mashup after embedded into Internship Mashup application.

Code snippet for the application:



```
<html>
<head></head>
<title>Internship placement in Selangor and Kuala Lumpur</title>
<body><iframe width="800" height="800" frameborder="0" scrolling="no" marginheight="0" marginwidth="0"
src="http://maps.google.com.my/maps/ms?
hl=en&amp;ie=UTF8&amp;source=embed&amp;msa=0&amp;msid=209829093773896666429.00049e1ddaa0af3197508&amp;
ll=3.074695,101.74713
1&amp;spn=1.097035,1.100006&amp;z=10&amp;output=embed"></iframe><br /><small>View <a
href="http://maps.google.com.my/maps/ms?
hl=en&amp;ie=UTF8&amp;source=embed&amp;msa=0&amp;msid=209829093773896666429.00049e1ddaa0af3197508&amp;
ll=3.074695,101.74713
1&amp;spn=1.097035,1.100006&amp;z=10" style="color:#0000FF;text-align:left">Internship Placement in
Selangor and Kuala
Lumpur</a> in a larger map</small>
</body>
</html>
```

Figure 4.10: Application Coding Snippet from Google Map

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 CONCLUSION

Mashup are certainly an exciting new genre of Web application. The combination of data and represent in into more visualize ways will attract and help user to use this system faster and more efficient.. In this paper creating Mashup using Yahoo Pipes is emphasized as it allow novice user like me to create one without having to write code or understanding programming concept. This paper also stressed on the challenges that commonly arise when creating a Mashup such as data extraction, data cleaning, data integration, service heterogeneity, and maintainability. Using Mashup to shared knowledge among students and staff is new and interactive way of communication using an information system. This will benefit the students who are undergoing for industrial internship to make relevant decision on choosing the best company that will suit them better. The MIP (Mashup for Internship Placement) can be easily embedded into other client application such as web browser and website for example into UTP e-learning system or student portal.

5.2 RECOMMENDATIONS

In terms of future work, more companies will be added into the database and will include companies that offer internship placement all over Malaysia including Sabah and Sarawak. This Mashup not only can be used for industrial training programme but also can be apply to other industry such as health, natural disaster, and higher learning education. For example in health research can be done to do Mashup to locate the countries that have higher number on certain disease. The proposed Mashup engine also can be used to locate country or city that recently was stroke by natural disaster like tsunami or earthquake. There will be research work to enhance several issues that may arise for example security issue in creating data Mashup, privacy issue and also the reliability issue on Mashups.

REFERENCES

1 Zang,N.,Rosson,MB.: 'What' in mashup? And why? Studying the perceptions of the web-active end users', (2008), The Pennyslavia State University.

2 Biörnstard,B., Pautasso, C : 'Let It Flow: Building MAshups with Data Processing Pipelines', ETH Zurich, Computer Science Department, University of Lugano, Faculty of Informatics.

3 Rattapoom, T., Pedro,S., Craig A.K,. : 'Building Mashups by Example', Jan(2008). Information Science Institute,Universtiy of Southern California.

4 Ort,E., Brydon,S., Basler,M. : 'Mashup Style Part 1: Server-side Mashups', May(2007). Can be access at: http://java.sun.com/developer/technicalArticles/J2EE/mashup_1/index.html

5 Richardson, L., Ruby,S. : 'RESTful Web Services.' O'Reilly, Sebastopol (2007)

6 Duane, M. : 'Mashups: The new breed of Web app', Aug(2006) :ibm.com/.../x-mashups

7 A Dennis, B H Wixom, D Tegarden : 'System Analysis and Design with UML Version 2.0 – An Object-Oriented Approach' J Wiley & Sons,Inc 2nd Edition, pg 11(2005).

8 Ennals, R., Brewer, E., Minos, G., Shadle, M., Prashant,G. : 'Intel Mash Maker: Join the Web', Dec (2007). SIGMOD Record.

9 Serge, A., Ohad, G., Tova, M., Neoklis, P. : 'MatchUp :Autocompletion for Mashups' June(2009). University of Calif Santa Cruz.

10 Hostip.info: hostip.info Website, http://www.hostip.info/

11 MaxMind LLC: GeoIP, http://www.maxmind.com/app/ip-location.

12 FraudLabs.com: IP2Location, http://www.fraudlabs.com/ip2location.aspx

13 IPligence Ltd : IPligence, http://www.ipligence.com/

14 Quova: Geopoint, http://www.quova.com

15 IP2Location.com: IP2Location, http://www.ip2location.com/

16 Sun Master Data Management Suite Primer :Data Mashup Engine, Dec(2008), SunMicrosystems, Inc. : docs.sun.com/app/docs/doc/820-5699/ref_mdm-primer-dm_c?l=en&a=view

17 Wikimapia , http://www.wikimapia.org

18 HealthMap.org : http://www.healthmap.org

19 Google Maps : http://maps.google.com.my

20 Google code : http://code.google.com

21 Geocommon : Google Maps for Heat Maps : http://geocommon.com

22 MIBAZAAR : blog.mibazaar.com

23 Enterprise Mashup : http://msdn.microsoft.com/en-us/library/bb906060.aspx

24 Yahoo Pipes: http://pipes.yahoo.com/pipes/docs?doc=modules

25 Yahoo Pipes: http://pipes.yahoo.com

26 Survey form online: https://spreadsheets0.google.com/viewform?formkey=dFFfWWRFZHE4YU9nMnpJYWFJNVo1Q2c6MQ#gid=0

27 Summary of survey responses : https://spreadsheets.google.com/gform?key=0AsFezEJq_y3DdFFfWWRFZHE4YU9nMnpJYWFJNVo1Q2c&hl=en&pli=1&gridId=0#cha

APPENDIX

APPENDIX 1 : SURVEY QUESTION AND SUMMARY OF RESPONSES

Task/Week	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14
Do a survey														
Gathering requirements														
Analyzing the problem statement														
Plan or design for software based														
Implementation /programming part														
Testing the system														
Maintenance.														

Gantt chart and work plan of the project semester 1

Task/Week	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14
Do a survey														
Gathering requirements														
Analyzing the problem														
Plan or design for software based														
Implementation /programming														
Testing the system														
Maintenance.														

Continue Gantt chart and work plan of the project semester 2.