METASEARCH ENGINE ON PROPERTIES

BY

Lim Jun Yuen

A PROPOSAL

SUBMITTED TO

UniversitiTunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONS)

Faculty of Information and Communication Technology

(Perak Campus)

JANUARY 2015

DECLARATION OF ORIGINALITY

I declare that this report entitled "Metasearch Engine On Properties" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature   :   _____

Name   :   Lim Jun Yuen

Date   :   21th March 2015

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude and appreciation to my supervisor, Dr. Alex Ooi Boon Yaik for not giving up on me and provide encouragement when doing this project.

I would also like to express my grateful thanks to all my friends and course mates for giving me motivation and inspiration when doing the final year project.

Last but not least, I would also like to thank my parents. They were always supporting me and encouraging me with their best wishes.

ABSTRACT

Buying a property is one of the biggest investments in life as it involves a lot of money. Besides that, it is a time consuming process when it comes to researching on desired properties and decision making. There are lot of available property listing websites that help user to search for property and provides property's information such as price, location and etc. However, property listed on the property website might not listed on other property websites. In order to get the best deal, user would require extra effort to access multiple property websites and compare the properties of the websites.

Moreover, most of the local property website only provides area based search and filters feature. There is no flexibility that allow user to search based on the geographical area scale of user interest. When user input "Kuala Lumpur" as the search query, the website will provide listed property that located within Kuala Lumpur. However, Kuala Lumpur is a huge geographical area that consist multiple smaller scale of suburb area. User might not familiar with the area of Kuala Lumpur and might not know how far apart between the area of interest and the location of property being stated as search result. User might take a lot of time to scale out search result that is not in the area of interest.

In this project, a web application with metasearch engine feature is developed to enable user to search property from several targeted property listing website in a single search. Multiple crawlers are developed to extract data from the property websites and aggregate to present as search result using the application. Besides that, the web application is implement with search feature that enable user to search property based on the area of interest that is drawn on the map.

LIST OF FIGURES

LIST OF ABBREVIATIONS

API   Application Interface Programming

AJAX   Asynchronous JavaScript

HTML   Hyper Text Markup Language

MYSQL   Structure Query Language

PHP   Hypertext: Preprocessor

TABLE OF CONTENTS

CHAPTER 5: SYSTEM TESTING

CHAPTER 6: CONCLUSION

Chapter 1: Introduction

1.1 Project Background

Buying a property is one of the biggest investments in life (Lee 2014). The reason is that buying a property involves a lot of money and it could affect individual's financial expenditure on other material. Moreover individual might need to consider whether they are financially affordable the property that they wanted. A statistic has shown that from year 2005 to 2012, housing loan itself stands 13% of total bankruptcy (Ching 2013). It is the third largest portion among all the total causes of bankruptcy (Ching 2013). Finally, there is a great potential that individual might suffer great lost from first-time buying a property provided that they do not do their research well before making the decision.

Doing research in getting information about the fair market value and home inspection will save up a lot of money for buyers (Fontinellen.d.).One of the ways to obtain fair market price is through Comparable Market Analysis (CMA) (Investpedian.d.). CMA is referring to an examination of prices which those similar properties that are being sold in the same area (Investpedian.d.). The examination is crucial in decision making because users and buyers could buy and sell properties based on fair market value. Besides doing research on fair market price, research on the area where the specific property is crucial factor to be considered before buying a house (Folgern.d.). Buyers and sellers are able to know how much to offer and how much to list for the price. A lot of research could be done to avoid from making bad decision.

There are a lot of ways that individual are able to gather the information to aid on their decision making. One of the ways is to look for local real estate appraisal for targeted property's information. Real estate appraiser is a state-licensed professional

who is equipped with knowledge about property market in their specialized area (Washington State Department of Licensing 2015). They are often hired by bank to determine the value of the property and decide how much to loan the borrower (Wickell 2015). However, it might take time to wait for the real estate appraiser to provide result that is requested by individual. Transacted price of the properties that are in the same area of targeted property is needed for comparable market analysis. One of way to obtain the data is by performing ad hoc search at JPPH which cost RM1 per search (JabatanPenilaiandanPerkhidmatanHarta2011). For users who have a desired property in mind, local real estate agents are able provide information about what are the property around the area of interest. However for a user who has not decided which area of property that they want to buy, local real estate agents are unable to provide much information as they only specialize on certain area. Moreover, the agent might persuade and influence user's decision making. What's more, with the advancement of technologies, there is a lot of information that people can get it from the internet for research purpose. There are lot of websites that provides properties information and property listing. However, most of the property website only provides property listing that helps user to search for property that they wanted and does not provide information to aid them on decision making. Moreover, user required extra effort to access multiple property websites and compare the properties from these property websites on their own. This is because the property data that is listed on one website does not provided on the other.

In order to optimize web searching on properties and provide property insight to aid on decision making. Metasearch engine on properties is proposed. A Metasearch engine is a search tool that uses other search engines data to produce their own results from the internet and utilizes the search engine that is provided from the website to obtain result (Metasearch Engine 2015).

1.2 Motivation and Problem Statement

Although there is lots of website has been created to provide property information and helps user to search houses that are for sale or for rent. Yet user still faces problems when referring such sources.

- It is a tedious process when user tries to search for house that best meet their requirement. There are huge numbers of property listed on the property website. Users are able to search through these listed properties by using the search tools and filters provided. However, user might not get the best deal if they only refer a single property website as property that is being listed on the website might not be listed on the other property website. User would need to note down property that they are being interested on a particular website and compare the search result with the other search result that is from the other property website.

- Most of the local real estate website only provides area based search and filters. There is a limitation as it provides search result with the scale of city and states. There are many search result within such huge scale of geographical area and these result are randomly displayed in a list format. For example, when user input "Kuala Lumpur" as the search result, the website will provide listed property that located within Kuala Lumpur. However, Kuala Lumpur is a huge geographical area that consist multiple smaller scale of suburb area. User might not familiar with the area of Kuala Lumpur and might not know how far apart between the area of interest and the location of property being stated as search result. User might take a lot of time to scale out search result that is not in the area of interest.

- Most of the local property websites only provides results about what are the properties within the search area and details about these properties. However, it does not provide much about the information of search area. The user unable to know whether the targeted property that is listed is overpriced due to the reason that user is unfamiliar with the price of property within the search area. User required to do research to find out what is the median price of the property within the area of interest if the property website does not provide. What's more it would require extra effort if user would wants to know the overall median price of the property that is summarized from multiple property websites.

1.3 Project Scope

This project aims to develop a web-based platform that provides search feature and enable comparison on property listed on multiple property websites. The proposed system will implement metasearch engine feature to help user search property from several property websites in a single search. Besides that, to ease user on making comparison between properties, the system would organize the extracted information and display the search result based on the search criteria of user.

In order to provide flexibility for user to search property that is within the area of interest. The proposed system includes drawing function to allow user to draw area of interest on the map and perform search based on the area that is being drawn by the user. Finally, the proposed system would also display median price and price trend that is calculated with charts format.

In this project, the system would only able to search property information from several local property websites and properties that are within the area of Malaysia. The system would first focus on allowing user to perform search on Selangor states and would expand the search area after the system functionality and feature is completely implemented and tested. The reason that Selangor area is being

focus first is that the capital city of Malaysia, Kuala Lumpur is located on the Selangor area and there's high number population of people in the area.

1.4. Project Objective

There are several objectives in this project and are listed below:

- **To optimize web searching on properties and reduce work load of user**

  The system will be able to represent user to search property information from multiple websites in a single search. The search result that is provided is being aggregated and collected from several websites and this reduce the need of user collect the information from different website on their own.

- **Increase the flexibility of searching properties**

  The system provides draw function to allow user to perform search unlike most of the local property website that only provide fixed area-based search and filters.

- **To provide information about the search area based on the data collected**

  The system would able to compute information about property that is within the search area that is either for sale or for rent. Information such as what is the median listing price, median rental price, number of different type of property and number of properties within price range will be computed. The information is computed based on the

Bachelor of Computer Science (HONS)
Faculty of Information and Communication Technology (Perak Campus), UTAR

data collected from multiple websites. Moreover, the information will be displayed with different charts format, such as pie chart, bar chart and line graph.

- **To enable user make comparison between properties from different websites**

  The system will provide search result on the properties that are being listed at different property website and is within the search area of user. The properties will be categorized with the type of the property and display along with the price of property to enable user make comparison on properties across from different websites. The system would also provide the source URL where does the information of property is being captured to allow user refer back to the source.

1.5 Impact, significance and contribution

Upon the completion of this system, user can search property information from several website with a single search. These could reduce the workload of user having to search property information across several property websites and a need to note down the search result for further comparison. Besides that, user able to make more inform decision as the system would provides information about the search area. Information such as the median listing price, price trend and comparison of price between comparable properties within the area allow user to have a general overview on what are the market price of the property. Moreover, the information could also help user that haven't decided on which place to stay as these information could help user to know what are the place of property that would meets the budget and requirement of user.

By implement drawing feature to the system, user able to draw the area of interest that they want to perform search on the map. These could scale out area of properties that user are not interested and perform search based only the area of interest of user unlike fixed area search feature that is provided from most of the property websites. Finally, most of the local property websites does not provide these functionalities and it does not include metasearch engine feature. The system able to help user have a better way on searching property information online and reduce the workload of research before decision making.

Chapter 2: Literature Review

Information is crucial and played a very important role in most of the decision making. Buying and selling property is one of the examples that required reliable, updated and accurate information to make a good decision. Therefore, the information that is provided by property websites is one of the criteria that need to be review as it shows how much does it able to help user on decision making. Moreover, the way of the websites presenting the information to the users is also crucial criteria. Websites that presenting information in a well and systematic manner could help user save up time because user are able to grasp the information easily that they needed. Finally, the usefulness of search tools in the property websites is also one of the criteria that must be review. A good search tool in the website could help user to search for information that they needed without browsing through the whole websites. Besides, it might also able to narrow down the search scope and provide more relevant result that meet the requirement for user. This could help user to save up lot of work and time as they need not to look through irrelevant results.

In this project, we will review 3 local property websites that are available on the internet which includes iproperty.com, propertyguru.com.my, propwall.com and 3 overseas property websites which include ziprealty.com, zillow.com and trulia.com. We will review the strength and weakness of these websites based on the review criteria mentioned before which are the information provided and search tools of the websites.

2.1 Websites reviewed according to their strength and weaknesses

2.1.1 iProperty.com

Like most of the property websites, iProperty provides details about the property that are being listed such as the price and location of the property. The website also provides the information regarding about what are the transportation, schools and other amenities nearby the targeted property. Information such as distance between amenities and targeted property are being provided as well as the location of these amenities is being shown through map. However, lots of inaccurate information in the websites. Firstly, some property details do not have the complete and accurate addresses about the property. This could cause confusion to the users and making it hard for them to locate the property on their own. Moreover, some of the property that is listed shows the wrong location of the property on map. An example of below figure proved that the targeted property located at Klang. However the map showed the property was located at Sekinchan which is 74km away from Klang.
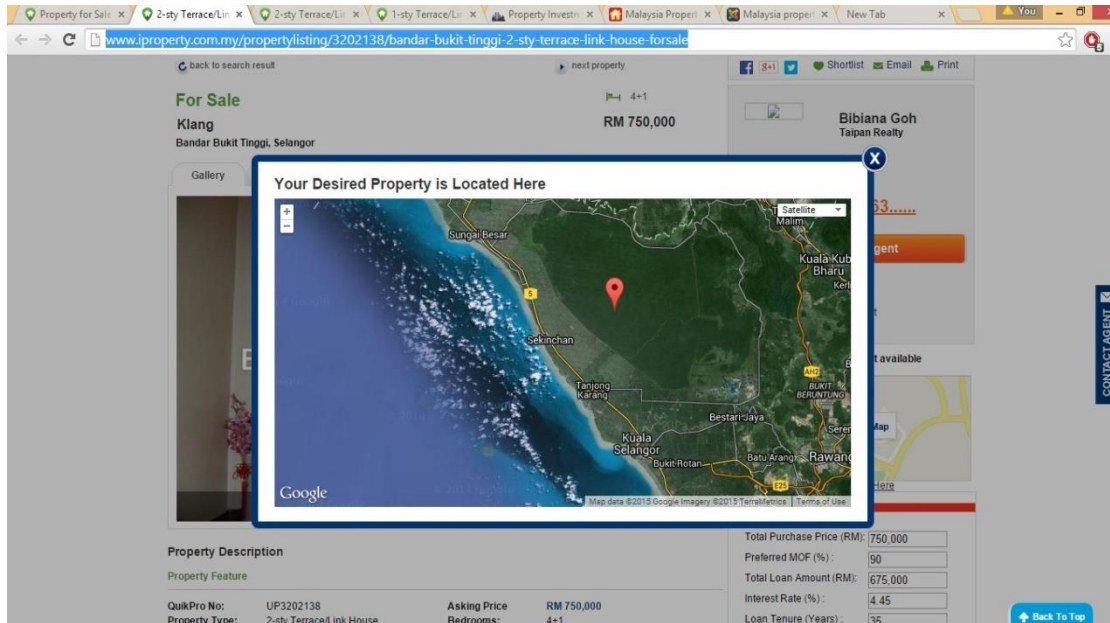
Figure 2.1: Location of property being shown

IProperty provides area search and filters to help user in searching the property that meet their requirement within a searched area. However, the scale of search area is being fixed by the scale of the suburb. There is a limitation as there are many properties within such huge scale of geographical area and it does not provide flexibility to scale out unrelated properties for users. User might find it difficult to look through all of these search result if the number of search result is huge. In other words, users are wasting time and energy in filtering unrelated information. Moreover, the websites does not provide much information about the searched area. Information such as average price of the property and number of sold property within the search area able to help user on making decision. Such information shows how much difference between prize of worth and comparable properties that are located at different area.

One additional search feature of iProperty over other local property website is that the website is able to search properties that are nearby to the current location of the user.

2.1.2 Propertyguru.com.my

Propertyguru provides most of the property details similar as iProperty.com. The website uses different icon to represent the type of amenities and location of these nearby amenities through map. This feature has advantage over iProperty. User could easily view what is the type of amenities that are surrounding the targeted property. However, the website have similar problem same as iProperty which is some of the property listed does not provide complete information. As example of figure below, one of the properties listed does not provide complete addresses and picture of the contact agent. Credibility of this information would be questioned due to seemingly untrustworthy agents who failed to post his own photo on his profile and provide incomplete property information.
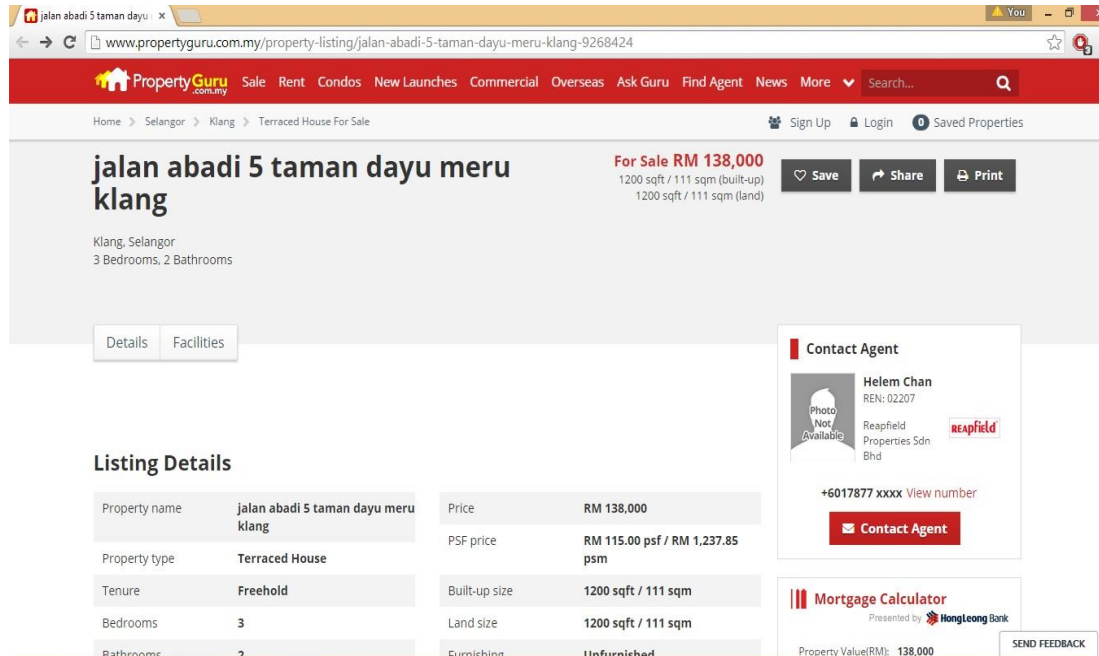
Figure 2.2: Incomplete property's information

As for the search tools, the website contain similar search tools with iProperty.com which are area based search and filters. Like iProperty, Propertyguru has a limitation whereby user is fixed to search the area with the scale of suburb or states. The website did not provide information regarding about the search area. However, the website did provide information regarding about the suburb, city and states of the country. This information could give a brief understanding about how helpful the area will be for people who are unfamiliar with the place. However, it does not help very much when making decision about buying property.

2.1.3 Propwall.com

Propwall has the similar search feature as other reviewed websites. These features include area-based search and filters. However, unlike other property website that provides a list of properties as the search result. Propwall would first return what are the suburb areas within the search area as search result. After that, user could choose the suburb area that they are interested from the search result and view what are the properties available for sale within the searched area. The strength of this feature is narrowing down and categorizes properties according to different smaller areas within a huge state. For example, when user search for properties within Klang area, Propwall narrowed down and categorizes all the available properties to smaller areas such as Padang Jawa, Bandar Bukit Raja, Bandar Bukit Tinggi and etc. This allow user to choose what area that they are interested instead of bombarding users with full lists of uncategorized properties within the huge area. The weakness of this feature is that user unable to instantly compare the differences between the properties in two different suburb areas. Users have to manually note down the information about selected properties and make comparison in order to choose which property to purchase.

Moreover, the websites also provide information and property details about the search area. Information such as types of properties and facilities are shown within the area. Moreover, the website also provides information about the average sales and rental price, difference between the asking price and transacted price of properties within the area. All of the information could help user to understand the market trend of the property within the area and evaluate the worth of the property. However, the websites does not provide analysis on every area of the property as most of the analysis is done and posted by the registered user.

2.1.4 Ziprealty

Ziprealty is one of the property websites that help users to search for property for sale. The website not only provides details about targeted property but also provides details about school that are within the search area. User able to search for the review and rating of the school of their preferred location and then search for properties on sale based on the school district. These could help user to save time on house searching if school district is an important factor that user need to be consider. Moreover, the website also provides information about the area which the property is located. The information includes median sales price trends, recently sold houses, and price history. Finally, the website also include house value estimator to help user to figure out what the value of listed property in today's market. The estimation is based on automated valuation model (AVM). AVM is a mathematically based computer software program that produces an estimate market value based market analysis of location, market condition, and real estate characteristics from information that is previously and separately collected (Standard on Automated Valuation Models (AVMs) 2003).

Ziprealty provides variety search tools compared to other property websites that had been reviewed. The website provides search tools that help users to find nearby comparable properties. Comparable properties are properties that share the similarities such as number of unit, size, age, distance and number of stories (Jacques n.d.). By simply adjust the value of range slider that act as search filters, comparables will be shown on the map as the search result and sort according to the distance that are nearest to the targeted property with list format. Moreover, the websites provide search by map feature that allow user to draw polygon on the map to indicate the search area. Available properties will appear within the area of the polygon drawn by users. This could provide flexibility to user and scale out location of unwanted

properties. Finally, the website also provide proximity search feature. User able to find properties that is within selected distance of preferred location.

2.1.5 Zillow

Zilllow has most of the feature that is similar to Ziprealty. The website also provides schools information that is nearby the targeted property. Moreover, the website also provide search by map feature which user could search properties based on drawing a polygon as a selection of area on the map. However, one of the advantages that Zillow has compared with Ziprealty is that the website also provides information of properties that are for rent. Not only the website provide home value estimation to help user in figuring out the selling price of the property in market, the website also provide rental estimation to determine what is the monthly rental price for a specific property. Zillow used Rent Zestimate, a propriety formula to estimate monthly rent price (Diane 2012). The result is compute by taking public data and similar rental listing data. Moreover, the number of rental listings in the search area could affect the accuracy of estimate value.

2.16 Trulia

Trulia is also one of the oversea property website that provides information of search area. However, the display format of Trulia is what made the difference between most of the property website reviewed. A heat map is a two-dimensional representation of data in which values are represented by colours (Margaret Trulia used heat map to display the area information that are surrounding to the property. Information such as the density level of crimes rates, levels of affordability, demographic and even the probability of natural disaster is being display through heat map. 2011). Heat maps allow users to understand and analyse complex data sets.
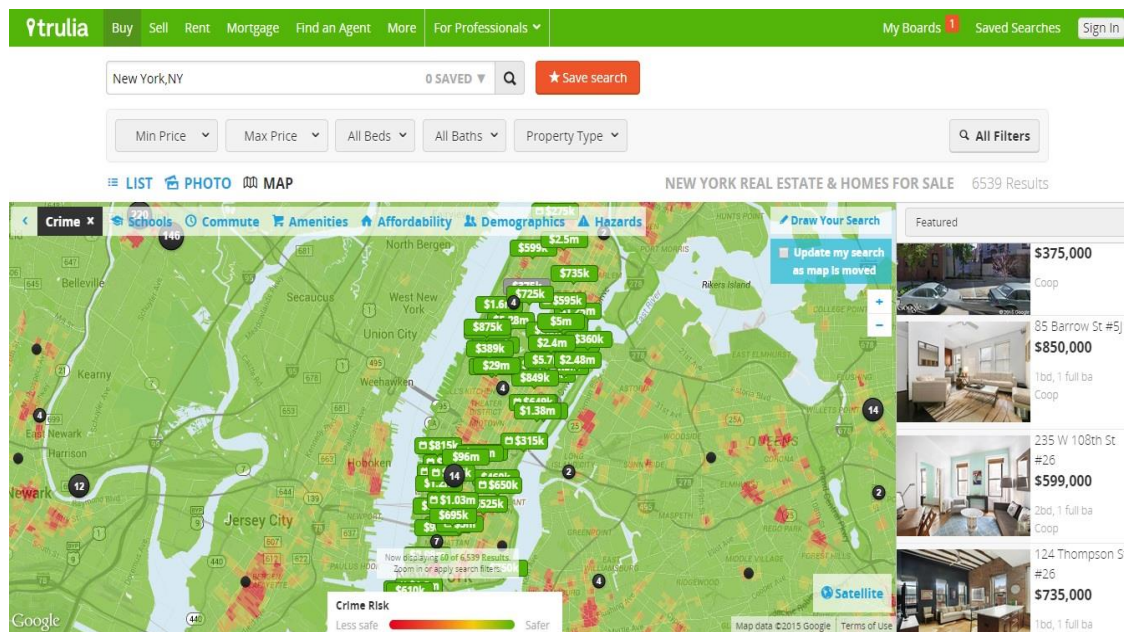

Figure 2.3: Heatmap of Trulia

2.2 Comparison Between Property Websites and Proposed System

	Search property by:			Information about area of interest	Comparing properties from several websites	Data from multiple listing websites
	Area or address	Refine search with filter	Drawing function			
Local						
Iproperty	Yes	Yes	No	No	No	No
Propertyguru	Yes	Yes	No	No	No	No
Propwall	Yes	Yes	No	Yes (not for all area)	No	No
Overseas						
Ziprealty	Yes	Yes	Yes	Yes	No	No
Trulia	Yes	Yes	Yes	Yes	No	No
Zillow	Yes	Yes	Yes	Yes	No	No
Proposed system	Yes	Yes	Yes	Yes	Yes	Yes

The table above shows the availability of feature from different property websites and compared with the proposed system. The differences and the similarities between each of the property websites based on the feature are clearly stated. All of

the property websites that have been reviewed does not able to make comparison between properties that are from other websites unlike the proposed system. Moreover, as clearly stated most of the property data that is provided by the website does not include the property data of the other website. This clearly shows that property website that has been reviewed does not share the same property listing with other website. Unlike with our proposed system which captured property information from several websites.

Therefore, the proposed system is quite different from the other property website as it contains additional feature compare with other property website.

Chapter 3: Methodology and Tools

3.1 Methodology

The software development methodology that employ in this project is evolutionary prototyping. This methodology is chosen because there is high level of user involvement from the start of development process of the system. Besides that, a working system is available early on the development stages. The prototype will be shown to the supervisor early to obtain feedback and evaluate whether it meet the specification of the system. After the evaluation, improvement and modification will be made to the prototype until the prototype completely evolves towards the final system.
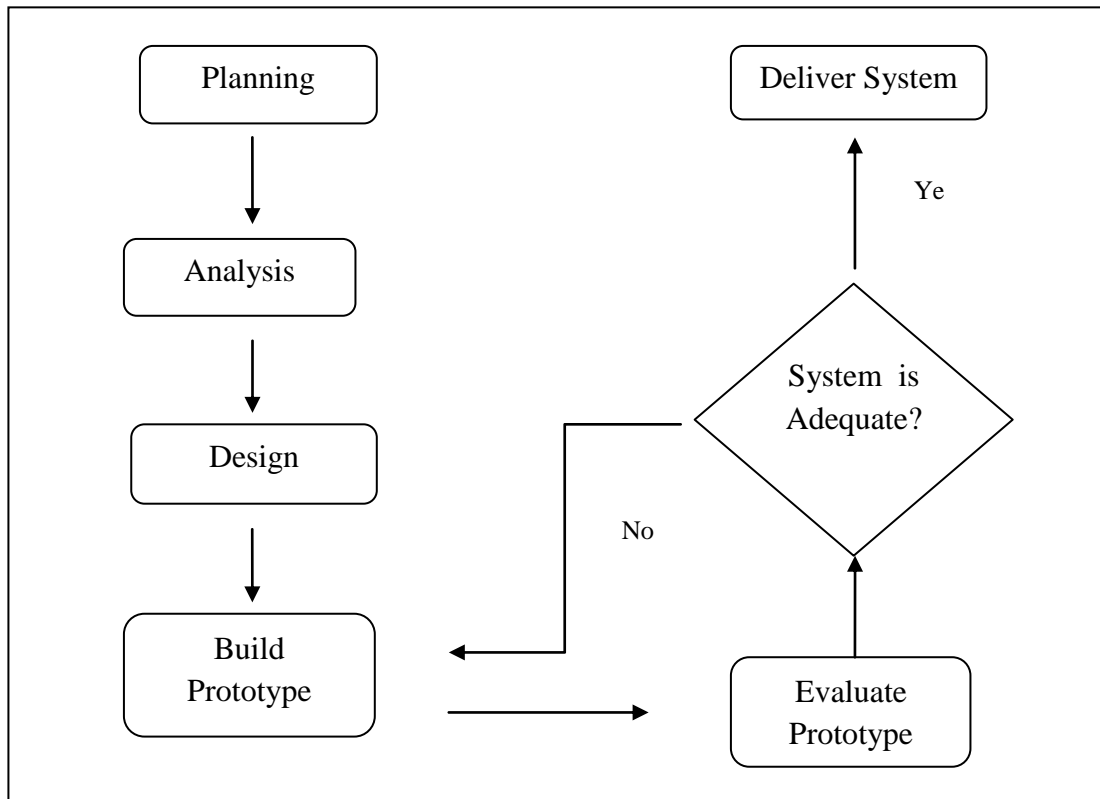
Figure 3.1 Evolutionary Prototype Model

3.2 Development Flow of Methodology

There are several phases in the development process of this project which include planning, analysis, design, implementation, evaluation and deliver of the final system. In this project, both planning, analysis and design phase will be undergoing in Project 1 while implementation and evaluation will be conduct during Project 2.

1. **Planning-** In this phase, the problem statement of the project is identified. Task analysis is conducted in this phase to understand on

how does the problem occur when user perform their tasks. Gantt chart is being used to schedule the time needed for the development of project and to check the milestones of the project.

2. **Analysis-** In this phase, study and analysis on the existing solution toward the problem is being conducted. There's no similar metasearch technology on properties being available online, therefore websites that provides property information that is frequently visit by user is studied. Moreover, the scope and objective of the project will also be defined to ensure that it could be developed completely within the timeline. Besides that, the requirement of the system is gathered and examine whether it could fulfill the need of users.

3. **Design-** In this phase, the interfaces, databases and system architecture is being designed. Operation flow of the system in terms of what are the data, process and steps involves with the system will be defined in this phased. Related diagram such as Use Case diagram, activity diagram and sequence diagram will be designed in this phase to gain insight on how does the system works. Study on the development tools and also acquired programming language that will be needed will also conduct in this phase.

4. **Implementation-** In this phase, a prototype with minimal requirement will be first developed. User interface will be developed to show the functionality of the system works. The prototype will be tested to find out whether it works as indicated and perform functionality that meet

the specification of the system. Technical problems such as error or bugs will be fixed in this phase if occur.

5. **Evaluation-** In this phase, user would evaluate the prototype developed to ensure that it meets the requirement and satisfaction of the user. Feedback and suggestion will be gathered to make further refinement on the system. User would also evaluate whether the system is adequate and does not require further improvement. If there's changes or further improvement need to be implement to the system than the system would need to repeat phase 3 and 4 to make adjustment.

6. **Deliver system-** In this phase, the final version of the system is being evolved and developed. The project will be demonstrated and presented in this phase.

3.3 Implementation Issue and Challenges

There are several challenges and implementation issue could be found in this project. First, in order to capture property information provided on the website. Inspection on the HTML of website is needed in order to search for the information that is on the webpage. However, the structures of HTML and the layout of property websites are different from each other. Therefore, programs that used to capture the information on the targeted property website might not work on the other websites. Extensive programming work would be needed to extract information from different websites. Secondly, the number of the properties data that is listed on a single websites is huge. What's more if the geographical search area of a user is huge, there will be huge number of data that needed to be collected and compute. There is a limitation capacity on the data that can be store in the database and high computation power will be needed to produce search result faster. Finally, there are inaccurate and incomplete information that can be found on the property websites. The number of inaccurate and incomplete information of property might affect the accuracy of the result and resulting undesired outcome.

3.4 Timeline

In this section, please refer to Appendix A to view the Gantt chart which shows the project timeline and tasks that need to be done during the development of the system.
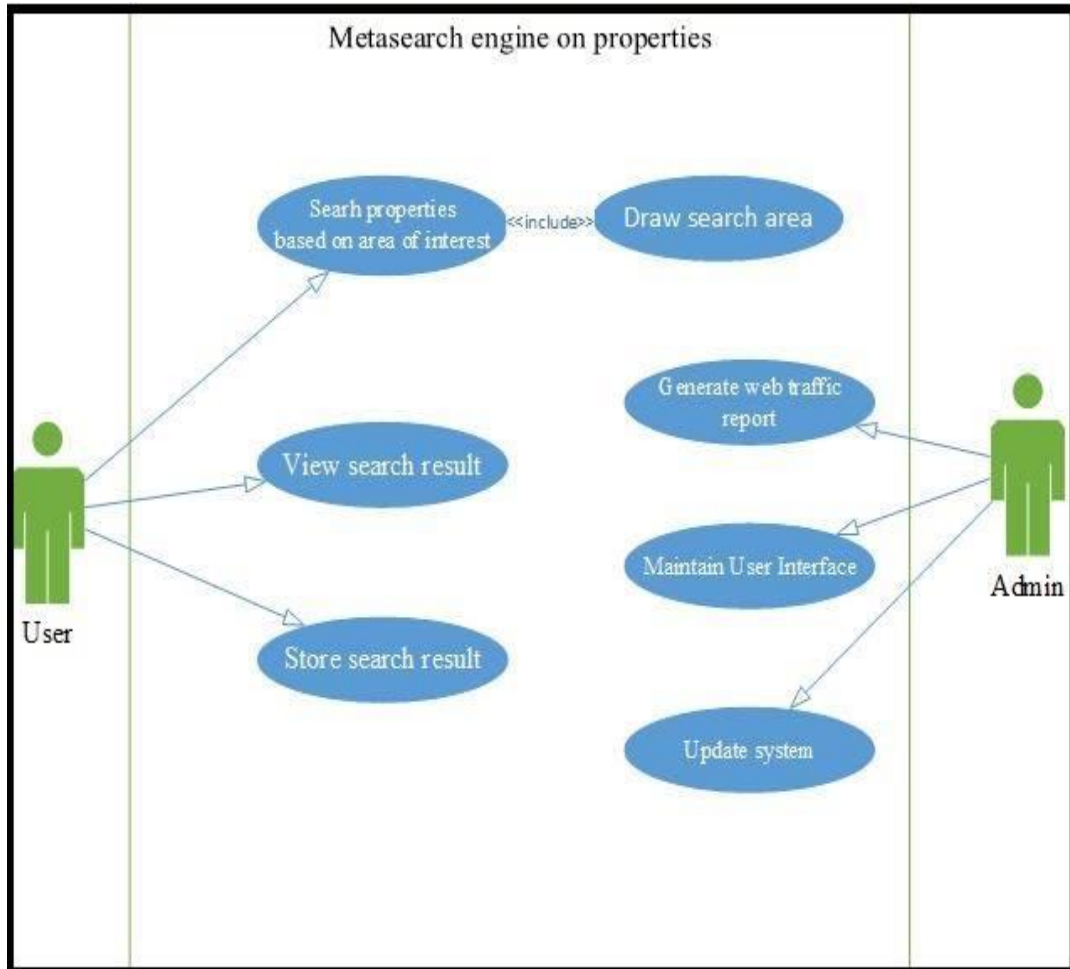
3.5 Use Case Diagram

Figure 3.2 Use Case Diagram

The use case diagram above shows the functionality of the system. It includes the type of user that interacts with the components of system. However, in this project the requirement of the system which we will focus on the functionality of system that will be utilize by user.

Based on the diagram above, user could perform three actions, which are searching properties based on area of interest, view search result and store search result. The user would need to draw search area first to indicate the area of interest that the properties located within. User could view the search result that shows properties that are within the search area. Besides that, information about search area such median listing price will also be display in the same time. Finally user could store the search result to refer back if needed.

3.6 Activity Diagram
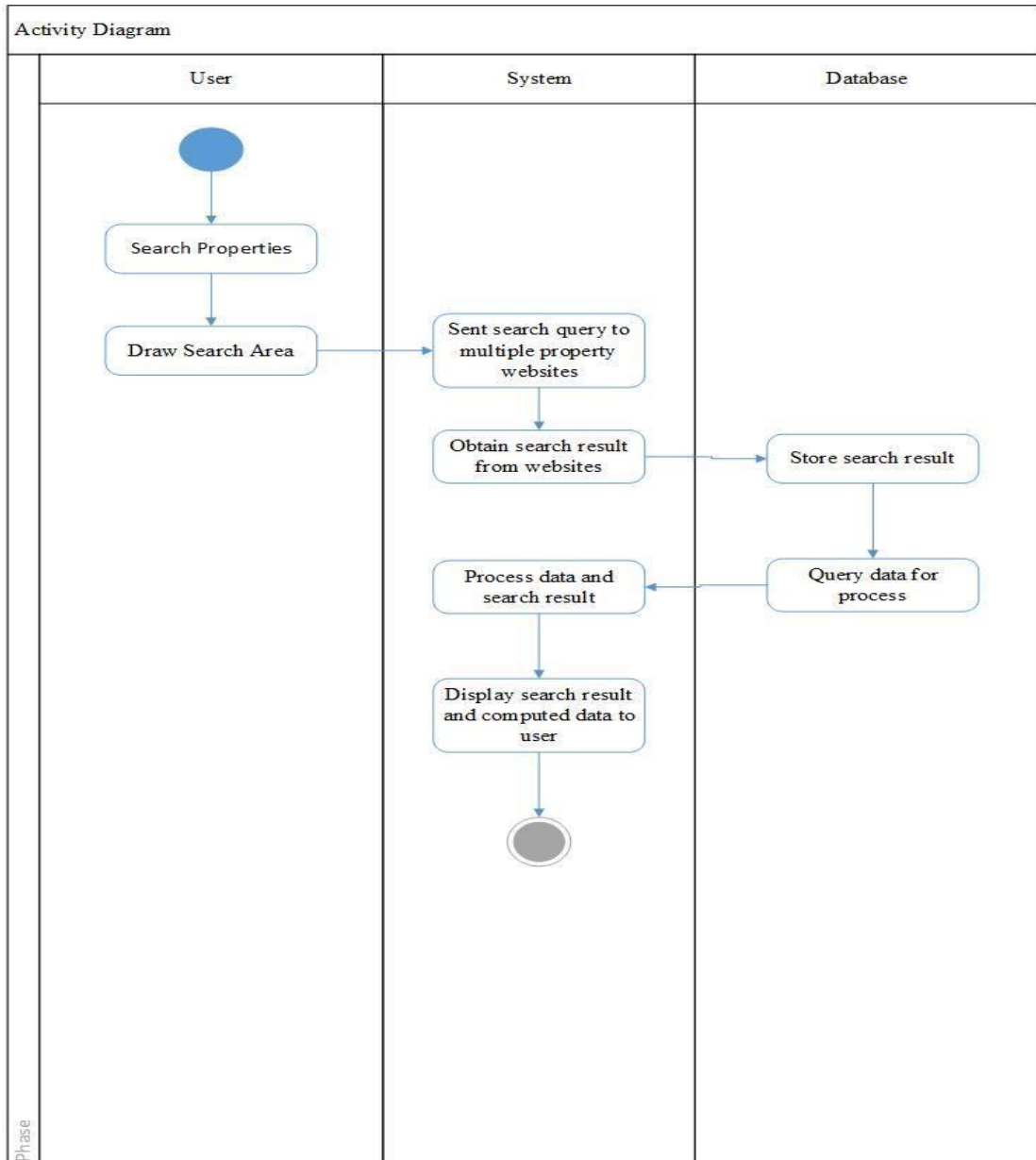
Figure 3.3 Activity Diagram

Figure 3.3 Activity Diagram shows the process flow of the system and the role performed by user, system and database.
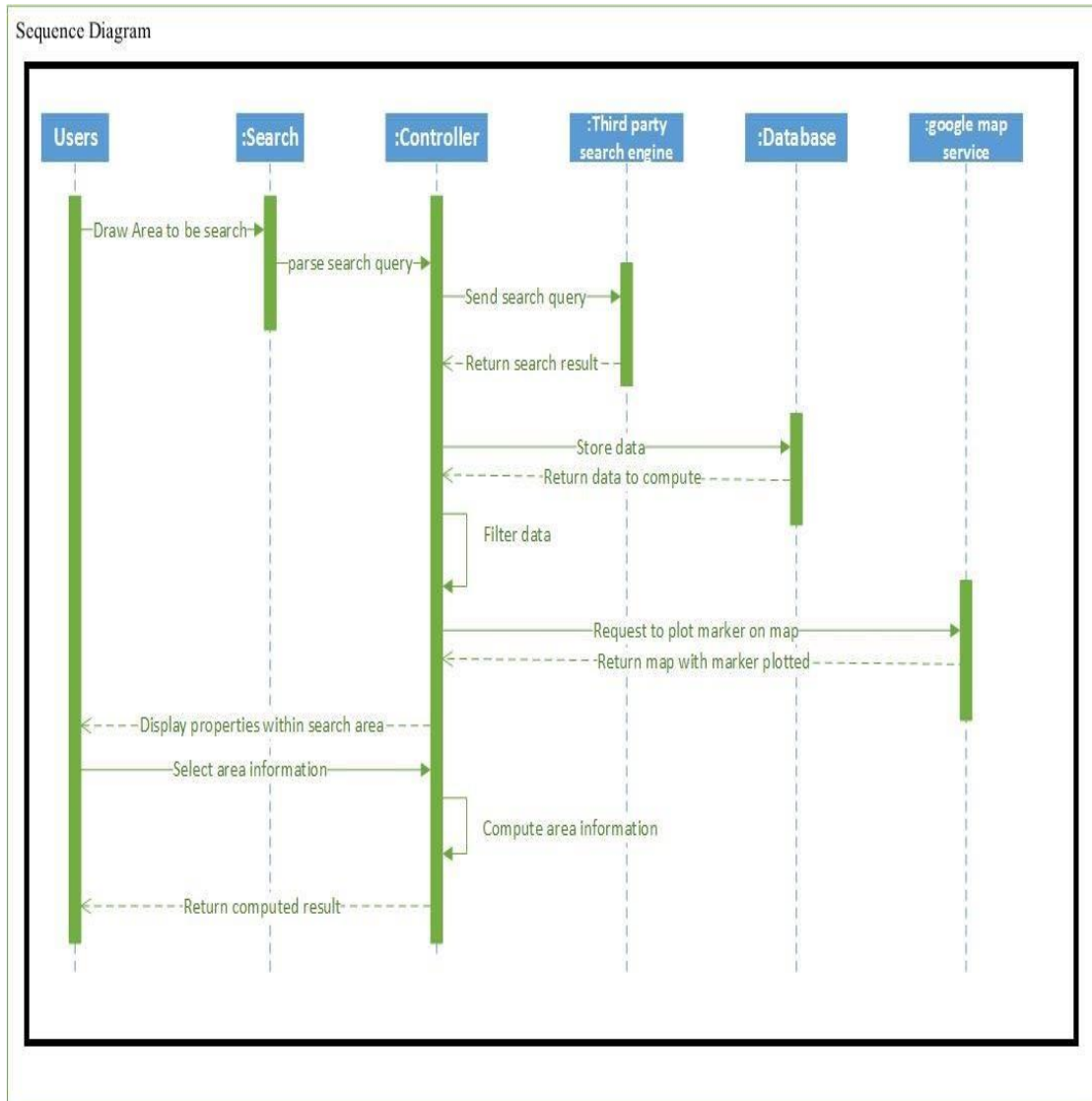
3.7 Sequence Diagram

Figure 3.4 Sequence Diagram

Based on Figure 3.4 sequence diagram describes what are the action that performed by objects at a specific time. Moreover, the diagram also includes the sequence of messages that are sent between the object that involve in the system operation flow.

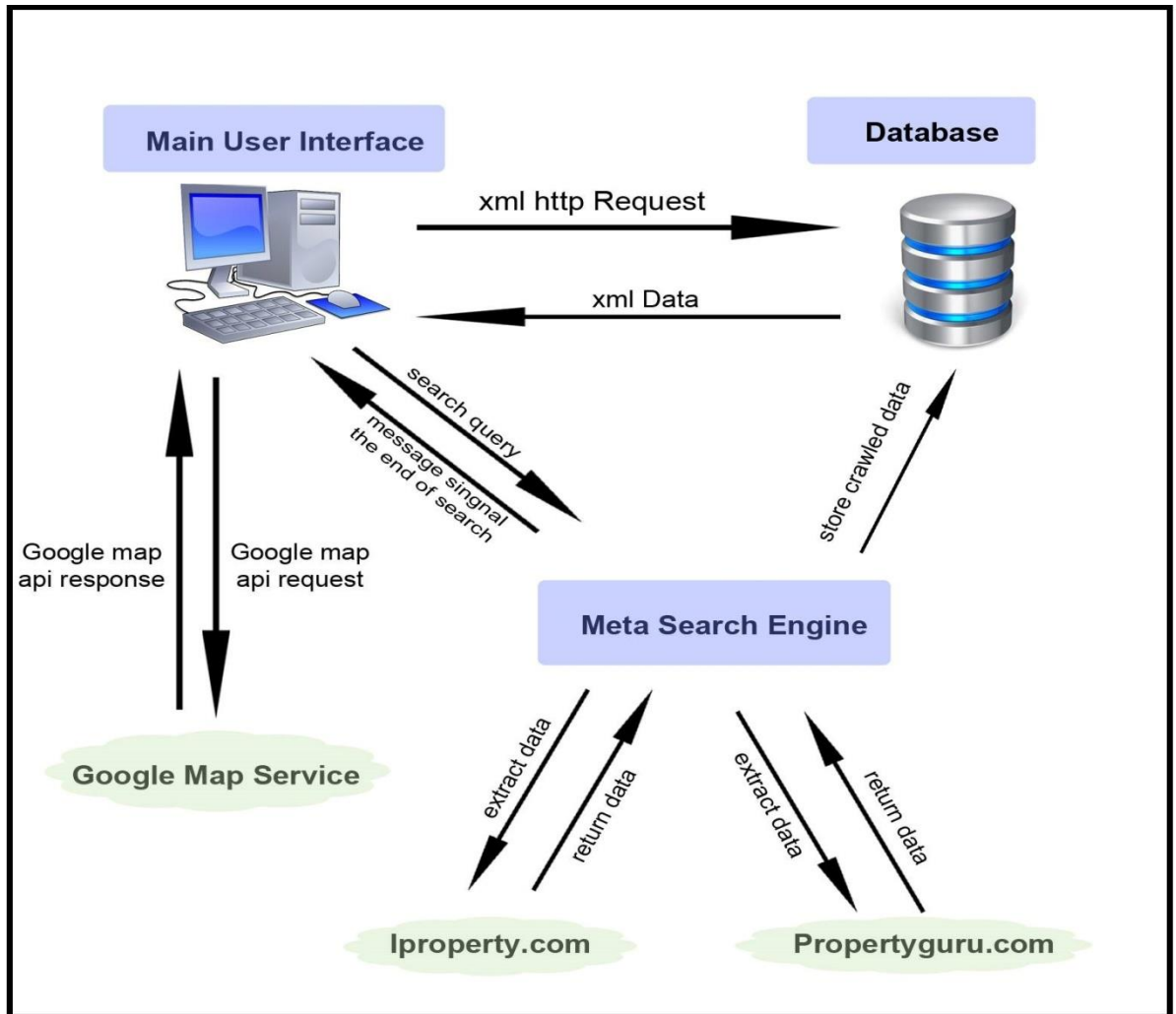## Chapter 4: System Implementation

4.1 System Architecture

Figure 4.1 Overview of system architecture

Figure 4.1 shows the overview on what is the flow of operation in the system. In the beginning, the main UI (User Interface) will send Api request to Google Map Service to load the Google Map Api. Google's Map and Google's drawing library will

be loaded to main UI after obtain response from the Google Server. The main UI will be connected to retrieve result data with XML format from the database by performing XMLHttpRequest. When user performs search action through the main UI, search query will be sent to the metasearch engine through the websocket. After metasearch engine received the search query from the main UI, search query will then pass to the crawler to extract property data from both "Iproperty.com" and "propertyguru.com". Data that extracted from the crawler will be store to the database. Once the metasearch engine fully crawled and extracted property data from the websites, a message will be sent to the main UI to indicate the end of search. After that, the main UI will then once again perform XMLHttp request to retrieve the crawled data from the database. These data will then be process and display as search result via the main UI.

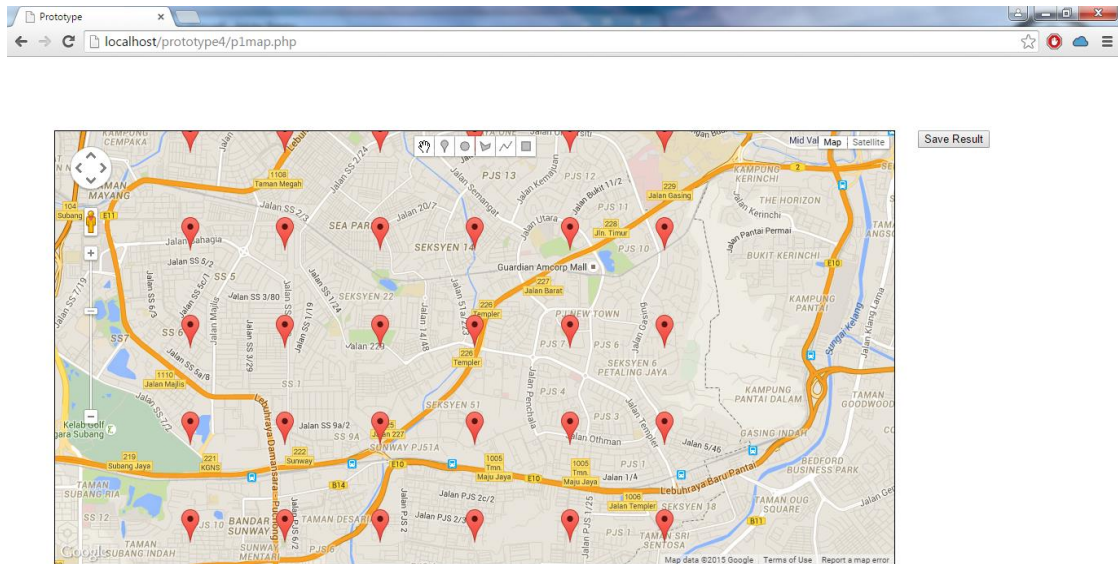**4.2 Main User Interface**

Figure 4.2 Main User Interface of the system

Figure 4.2 shows the main user interface of the system that user use to interact with the system and perform search action. The main user interface of the system is included with Google's maps which act as a platform that allow user to draw the search area that they want to search. The red marker is plotted onto the map along with the map in the beginning. These red markers are plotted onto the map in order to indicate what is the addresses and area that the marker located.
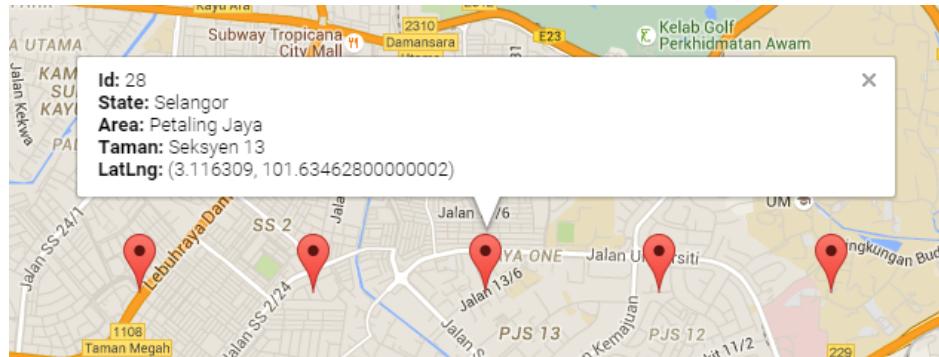
Figure 4.3 Information displayed by red marker

Figure 4.3 shows that once user clicked on the red marker, information such as what are the state, area, taman, position of the red marker located will be display. The source of the information where the marker located is obtain from "maps.google.com" and is recorded to database.

4.2.1 Drawing Function

This function is implemented by using google maps drawing libraries. Once user have complete drawn the polygon which is the search area on the map, drawing manager event listener will be trigger. By utilizing google.maps.geometry.poly.containsLocation function, the red marker that included inside the polygon will be verified.

```
//drawingManager eventlistener
google.maps.event.addDomListener(drawingManager,"polygoncomplete",function(polygon){
  var containState=[];
  var containTaman=[];
  var containArea=[];
//market that are within the polygon drawn will be stored
  var containMarker=[]
  var areaID=numPolygonCreated++;

//find what are the marker inside the polygon
  for(var i=0;i<markerArray.length;i++){
    if(google.maps.geometry.poly.containsLocation(markerArray[i].position,polygon)){
    containMarker.push(markerArray[i]);}
  }
}
```

Figure 4.2.1.1 Sample code for event listener and function that find the red marker inside the polygon.
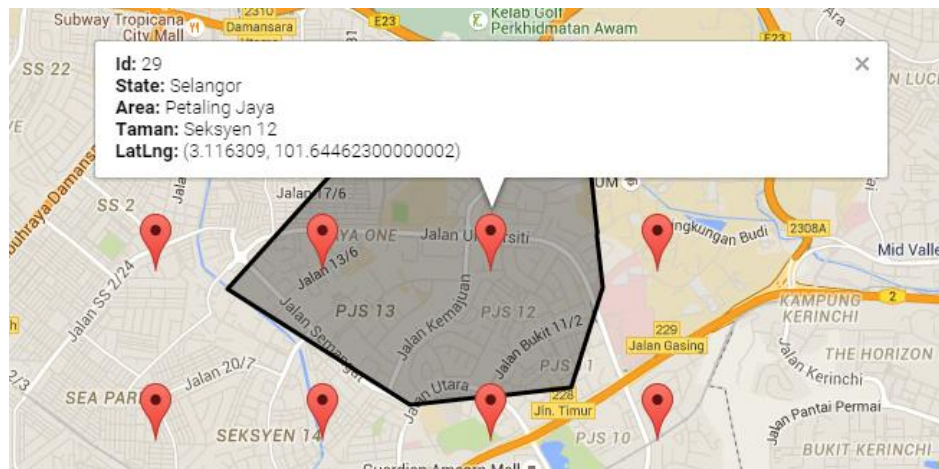

Figure 4.2.1.2 Red markers that are included inside the polygon drawn

Figure 4.2.1 shows that the red markers that included inside the polygon drawn by the user are marker that located on Seksyen 12 and Seksyen 13. Hence, the system would estimate that the search area drawn by the user are included with these 2 taman area by verify the marker that included inside the polygon drawn.
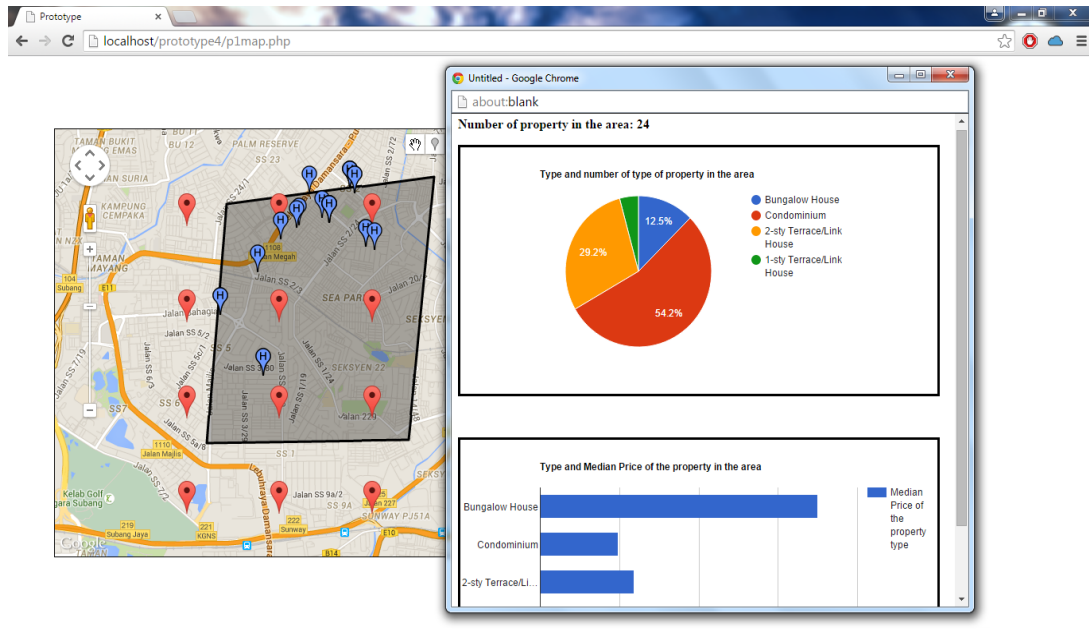
4.2.2 Display Result

Figure 4.2.2.1 Result displayed after user performed search.

Figure 4.2.2.1 shows the search result that display to the user. Blue markers that plotted onto the map are the result of property that are within the search area of the user. Information about addresses, latitude, longitude, property type, asking price, crawled date and URL will be display once the user clicks these blue markers. User can also view the source information that displayed from the blue marker by inserting URL that is shown to the web browser. Besides that, area's information regarding about what is the number of property in the area is calculated and display as shown in Figure 4.2.2.1. Number of the type of property in the area is display with pie chart and median price of the type of property in the area is displayed with bar chart. These

information is calculated based on the aggregated result from both iproperty.com and propertyguru.com

4.3 Database

Database is used to act as storage media in the system. In this project, MySQL is being implemented as database due to the reason that it is compatible with both PHP and Python programming language. The database consists of three tables which are "gridmarker", "crawlmemory" and "crawledrecord" table. "gridmarker" table is used to store the information about the area of location which will be used to plot red marker. Information such as marker_ID,state, suburb, taman, latitude and longitude will be stored in this table. "crawlmemory" and "crawlrecord" table consist the same field which are ID, Addr1, Addr2, Latitude, Longitude, PropType, AskingPrice, CrawlDate and URL. However, the usage of these two tables is different, "crawlmemory" table is used to store search result from the metasearch engine temporary. Once user decided to exit the system and not to save the searched data, the data that stored in the "crawledmemory" table will be deleted. While, "crawledrecord" table is used to store search result that user decided to store permanently for future reference.

4.4 Metasearch engine

The purpose of implement metasearch engine is to able user to access different websites and get the aggregated data from these websites in a single search. In this project, the property's website that access by the system is iproperty.com and propertyguru.com.my. These websites are chosen due to the reason that they contain huge number of property listing in Malaysia and provide property information such as price, location, property type and many more.

Google Chrome browser is used to study the HTML page structure of each website and inspect web elements that are needed to extract by the crawler. Program called crawler is developed to crawl and extract the data from these respectively websites. In this project, different strategy and method is implemented by the crawler to crawl the targeted property website due to the reason that each property websites has its own HTML page structure. Below shows the process of the crawler as well as the method and strategy implemented.

4.4.1 Crawler of iproperty.com

Most of the URL of website contains query string. For example "http://example.com/w/index.php?field1=value1&field2=value2" is a URL that contain query string of "field1=value1&field2=value2". The query string of URL will be passed to the query program to retrieve information or lead the user to another webpage's content. For iproperty.com, the URL of query program that is link to the search function of the website is http://www.iproperty.com.my/property/searchresult.aspx?. By manipulating the "k=" parameter in the query string of URL, the crawler able to access to the search result page same as user insert the keyword that they wanted to search on the website.

```
def extractIproperty(query,db):
    #process querytxt with keyvalue
    querytxt=query.replace(' ','+')
    #webpages of iproperty
    seedurl="http://www.iproperty.com.my/property/searchresult.aspx?t=S&gpt=AR&st=SE&ct=Petaling+Jaya&k="+querytxt+"&pg="
```

Figure 4.4.1.1 Piece of code that shows the URL of search result page

As Figure 4.4.1.1 shown, the string query with
"t=S&gpt=AR&st=SE&ct=Petaling+Jaya" is set to access search result page that is
property type of all residential, located in Selangor state and Petaling Jaya city. After
obtain the query message, the system will then insert the query message to the string
query of parameter "k" in the URL to access the search result page.

In this project, urllib2, a python module, is being used to fetch the URL of
search result page. The source code of the search result page will be obtained by using
urllib2.urlopen function. After that, the crawler will then extract the data from the
result page by matching the regex (regular expression) that navigate to the content
needed to extract with the text of source code of the result page. The crawler will first
extract the total number of result page that listed on the first search result page to
know how many pages that the crawler needed to be crawl through. After that, the
crawler will then extract the URL of property listing that contains in the search result
page. By accessing the URL of property listing that listed on the search result page,
data such as asking price, addresses, property type, longitude and latitude of property
will be extracted using the same method of matching the source code of the property
listing page with regex. Data that extracted from the property listing page will then be
stored to the database. Figure 4.4.1.2 shows the regex that defined to extract the data
of the property listing page.

```
#extract the property data of the webpage
def extractbranchIproperty(url):
    addr1='"title_text.*?">([\S\s]*?)<'
    addr2='"text_post.*?">([\S\s]*?)<'
    propertyType='id=pdPropertyTypeDescription>([\S\s^<]*?)<'
    askingprice='id="pdAskingPrice.*?">([\S\s^<]*?)<'
    longlat='google.maps.LatLng\(([\S\s^<]*?)\)'
    crawlDate=time.strftime("%x")

    #Fetch URL and obtain the HTML code of search result page
    htmlfile=urllib2.urlopen(url)
    htmltxt=htmlfile.read()
```

Figure 4.4.1.2 Regex and urllib2 used in the crawler.

After finish accessed the URL of property listing that contain in the first search result page, the crawler access to the next search result page by manipulating the "pg=" parameter in the query string of URL. Same process will continue until the data of property listing in the last search result page have been full extracted.

4.4.2 Crawler of propertyguru.com

Unlike the crawler of iproperty.com, problem occurs when the crawler of propertyguru.com attempt to extract the data from the website. Below are the methods that are implemented by the crawler of propertyguru.com and problems that arise during the implementation of the method.

4.4.2.1 Urllib2

The crawler of propertyguru.com is able to extract the data using similar method which the crawler of ipropertyguru.com implemented in the beginning. Urllib2 is used to obtain the source code of result page and extract the data from it.

However, problems arise after the layout of the propertyguru.com website undergoes updates. The main problem after the updates is that the source code that obtained using urllib2.open function does not correspond to the source code that inspects using Google Chrome Browser anymore. The content that needed to extract are not found on the source code of result page that obtained using urllib2.open function. However, the source codes that inspect using Google Chrome Browser contain the data that are needed to extract.

4.4.2.2 Mechanize Browser

Mechanize browser is a headless browser, which does not contain graphical user interface and is used for test automation in web application. In order to solve the problem and obtain the similar source code that inspects using Google Chrome Browser, Mechanize browser is implemented in the crawler.

```
def setBrowser():
  br=mechanize.Browser()
  #set cookies handler
  cookiejar=cookielib.LWPCookieJar()
  br.set_cookiejar(cookiejar)

  #ignore robot.txx
  br.set_handle_robots(False)
  return br

def ipropertyUrlopen(br):
  base_url="http://www.propertyguru.com.my/property-for-sale/1?market=residential&freetext=selangor&limit=10&sort=date&order=desc"

  #get the source code of the resultpage
  urltxt=getPageTxt(base_url,br)
  return  urltxt

def getPageTxt(url,br):
  response=br.open(url,timeout=60.0)
  urltxt=response.read()
  return  urltxt

def main():
  #set browser
  br=setBrowser()

  #open the main page
  urltxt=ipropertyUrlopen(br)
```

Figure 4.4.2.2.1 Piece of code of the crawler that accesses the URL of the propertyguru.com using Mechanize Browser.

As shown in Figure 4.4.2.2.1, Mechanize browser is first set up before use to open the webpage of propertyguru.com. Mechanize browser is set to handle cookie automatically and ignore robot.txt when used to access the webpage of propertyguru.com.

In the first attempt of using Mechanize browser, the browser unable to access the webpage of propertyguru.com and receive a HTTP 405 response from the server as shown in Figure 4.4.2.2.2. This happened due to reason that the website does not welcome bots or program to visit and able to identify the crawler as a bot.

Bachelor of Computer Science (HONS)
Faculty of Information and Communication Technology (Perak Campus), UTAR

Figure 4.4.2.2.2 Result of implementing Mechanize browser to access the webpage of propertyguru.com.

When user visits a webpage using a browser, the browser would send a header to the server of the website. The header contain a user-agent string which provide information about what are the browser that are using to make request to the server, its version number and details about the system. In the second attempt of accessing the webpage of propertyguru.com, mechanize browser is set to add header of user-agent string that contain similar user-agent string as a Google Chrome Browser. The user-agent string is modified in order to reduce the suspicious of server to identify the program as bots and imitate as a real user visit the webpage using Google Chrome Browser.
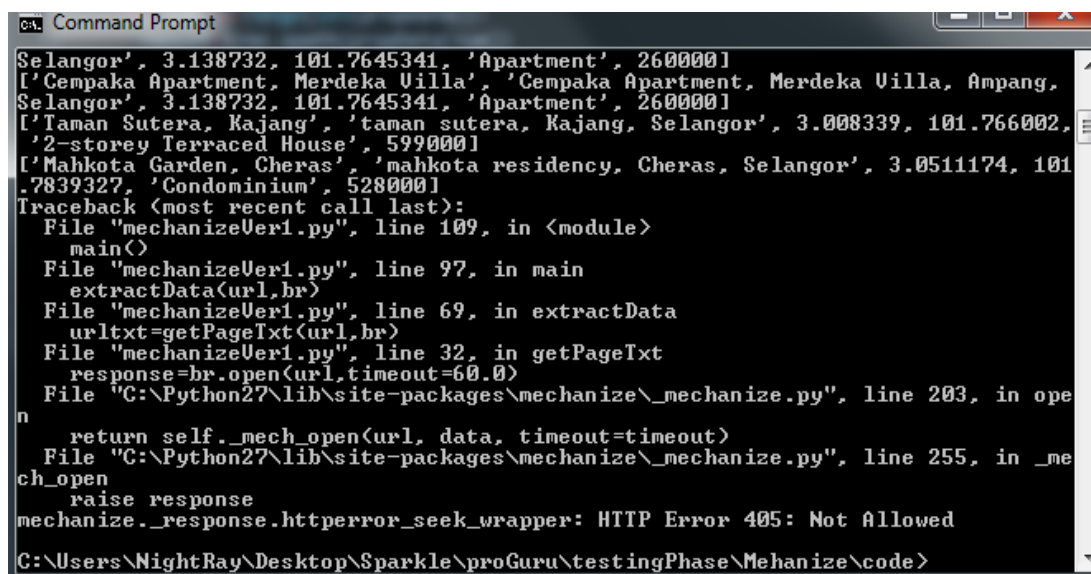

Figure 4.4.2.2.3 Piece of code that set the user-agent string of header.

Besides that, the speed of crawler crawling to the next result page is also controlled by implementing pause function to the crawler. Before the crawler visit the next result page, pause function will be executed to wait for 3 to 15 second. The function is included in order to imitate the behavior of user that browser through the website. A real user would need to take time to look through the content of the webpage before visit to the next webpage unlike a program.

After modified the header of the browser and include pause function to the crawler, the crawler is executed again as an attempt to access the webpage of the propertyguru.com. The crawler is finally able to access the webpage of propertyguru.com and extract data of properties listing from the website. However, the crawler unable to continue after extract four property listing data and receive HTTP 405 response again as shown in Figure 4.4.2.2.4.

Figure 4.4.2.2.4 Result after modified the header of browser and included pause function to the crawler

Bachelor of Computer Science (HONS)
Faculty of Information and Communication Technology (Perak Campus), UTAR

4.4.2.3 Selenium WebDriver

Selenium WebDriver is a tool used to automate browser for web application testing purposes. The difference between Mechanize and Selenium is that Mechanize is a headless browser that accesses the HTML of the webpage directly. While Selenium control browser application such as Firefox or Google Chrome to access the HTML of the webpage.

In this project, Selenium WebDriver is being used to automate Firefox browser and act as a crawler to crawl the webpage and extract the property data from propertyguru.com. Selenium provides find_element_by_xpath method which is a method that locate web's element in the webpage using XPath. By using the method provided, property data in the webpage is located and extracted from the webpage. Besides that, Selenium also able to automate low level interaction such as mouse movements, mouse button actions or key press on browser.

```
#wait the element to be loaded on the browser first and locate the element
searchBox=wait.until(EC.presence_of_element_located((By.NAME,'freetext')))
#send searh query to the searchbox
searchBox.send_keys("Petaling Jaya, "+queryChosenTaman[queryCount])
#send enter enter key
searchBox.send_keys(Keys.ENTER)
```

Figure 4.4.2.3.1 Piece of code on how Selenium is being used to interact with the element in the webpage

The code shown on Figure 4.4.2.3.1 is an example on how Selenium is being used to automate browser and perform search action on the webpage of propertyguru.com. First, element will be waited to load on the browser. The element that will be located in this case is an element with name of "freetext", which is the

search box on the webpage where user input their search query to perform search. After the element located, query text will then be send to the searchbox using searchBox.send_keys(querytext) function. The search action will then be perform after searchBox.send_keys(Keys.Enter) is executed which equivalent as a user press on the "Enter" Button on their keyboard.

By using browser automation tool, property's data from the propertyguru.com is able to be extracted. However, the automation process unable to continue when propertyguru.com redirect the browser to Captcha page as shown in Figure 4.4.2.3.2.
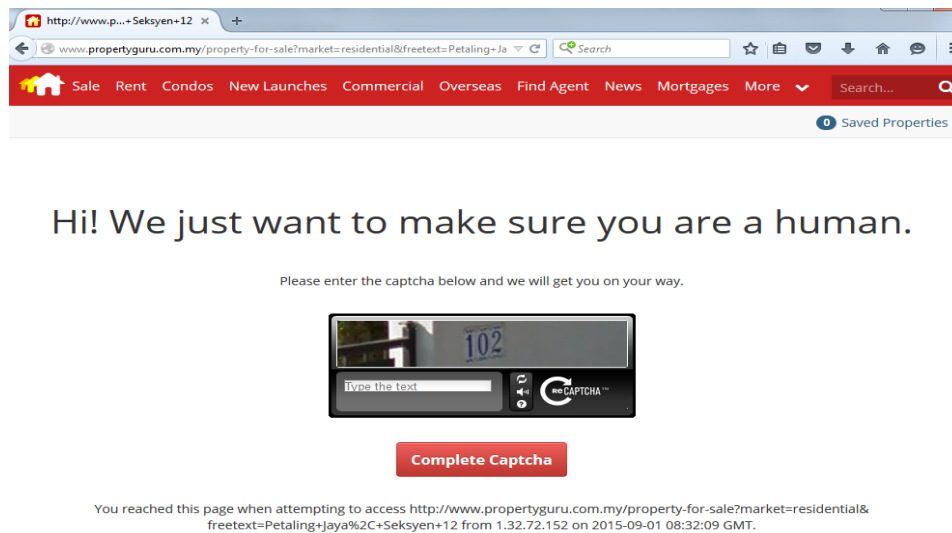

Figure 4.4.2.3.2 Captcha page.

The solution to overcome such problem is that user would need to manually enter the captcha text when the captcha page appears. By set up the browser to wait for 30 second before proceeding to the next page, the process will be continue after user enter the captcha text within the 30 second.

Chapter 5: System Testing

System testing is an essential process in order to makes sure that there are no error on the project before it is delivered to the end user. In this project, two test cases were conducted to test whether the system does what was intended.

5.1 Test Case-1

In the first test case, the crawler of the system is tested to check whether it extract all of the property listing data from the website that is requested by the user. In this test, only the crawler of iproperty.com will be tested. This is due to the reason that the crawler of propertyguru.com uses browser automation method to extract the data from the website. Hence, the data that has been extracted is being display with the browser when the crawling process is running along.

First, we perform search using the system by drawing a polygon that is within the area of Seksyen 12 and Seksyen 13 in the city of Petaling Jaya as shown in Figure 5.1.1.
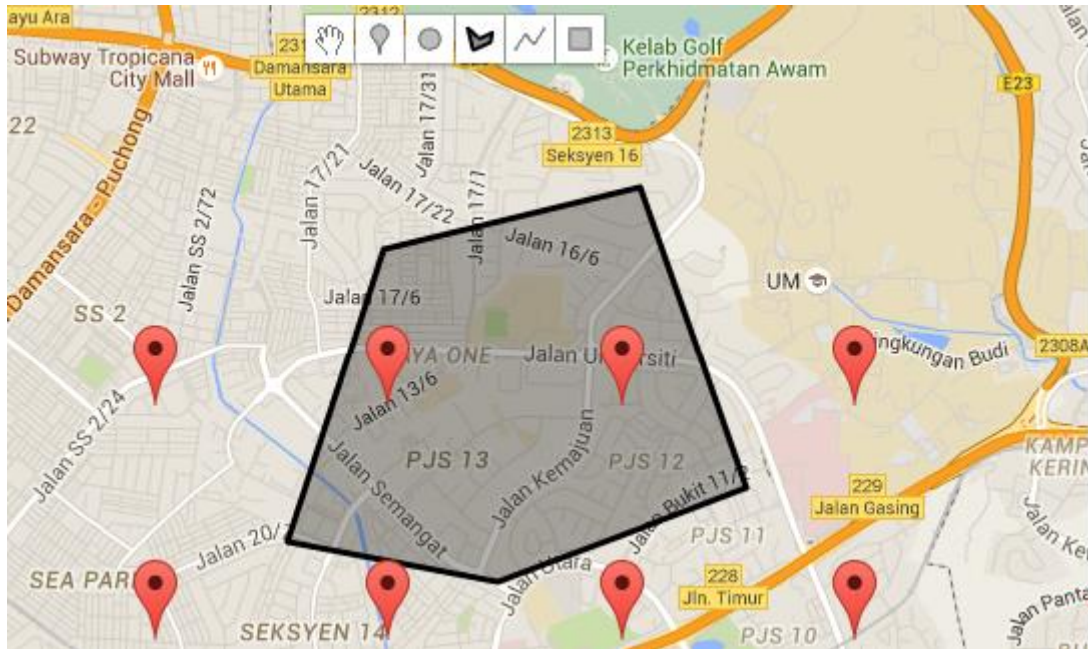
Figure 5.1.1 Polygon drawn which include the area of Seksyen 12 and Seksyen 13

After that, we perform the second search by inserting the same search query which is Seksyen 12 and Seksyen 13 using the website of ipropertyguru.com. Finally, we make comparison on the search result and check whether the number of result extracted by the crawler is similar to the result from the websites.

Figure 5.1.2 Result of the number of property listing that have been extracted

Bachelor of Computer Science (HONS)
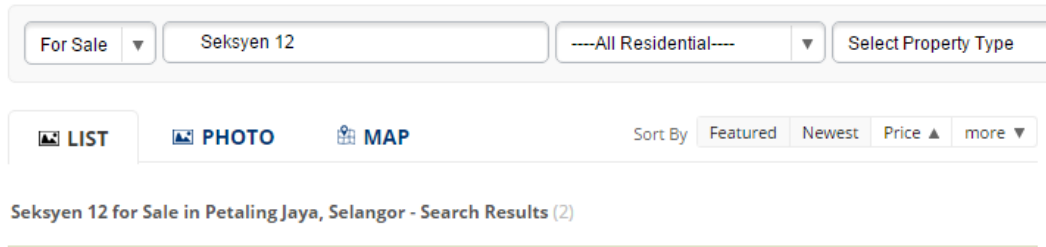Faculty of Information and Communication Technology (Perak Campus), UTAR

Figure 5.1.3 Number of search result for Seksyen 12 using the website
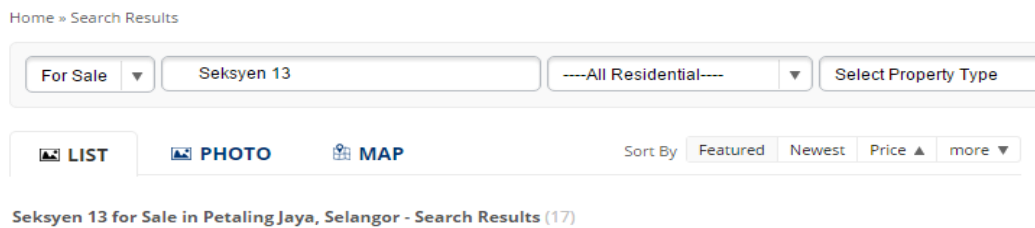
Figure 5.1.4 Number of search result for Seksyen 13 using the website

Based on the result as shown in Figures above, the total number of property listing that extracted by the crawler is 18. While the total number of search result that shown by the website is 19. The difference of the number of result shows that the crawler unable to fully extracted property listing's data from the website. This is due to the reason that the URL of one of the property listing page is different among the others. The property listing page that unlike others is with the URL of http://newlaunch.iproperty.com.my/... The crawler haven't include this type of URL to be crawl yet.

5.2 Test Case-2

In this test case, the system is tested to check whether the result that display after the search is accurate. First, we perform a search by drawing a polygon. After that, search result of the property listing's data that is located within the polygon will be displayed. Besides that, the system would provide information regarding property that located within the area such as the total number of property, number and type of the property, and median price of the type of property. We then manually calculate and check whether the result that provided by the system is similar to what we have calculated.

Figure 5.2.1 Result of the property listing that is within the search area

Number of property in the area: 5

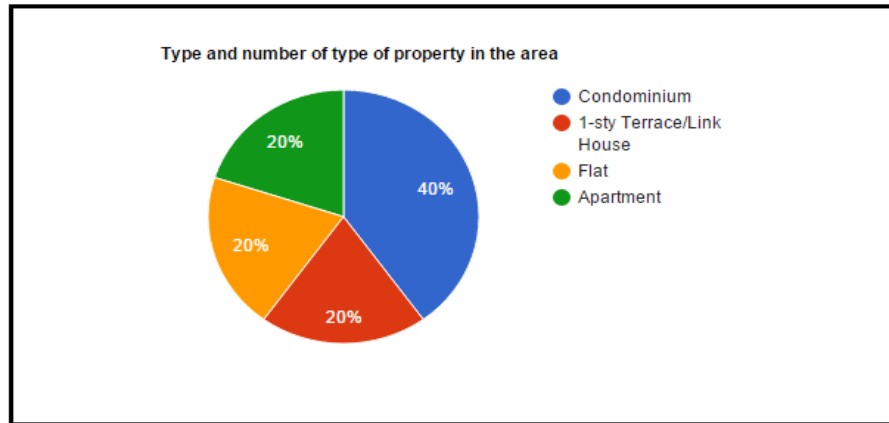**Figure 5.2.2 Result of the total number of property and type of the property in the area**
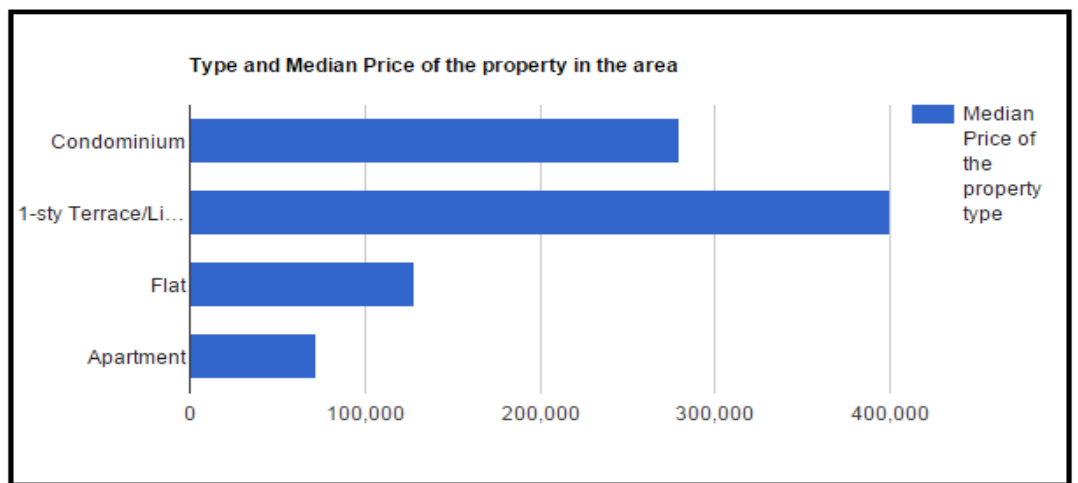
Figure 5.2.3 Result of the type and median price of the area

Figures above shows the search result that provided by the system. As shown in Figure 5.2.1, there are total of five plotted blue marker on the map which shows the

Bachelor of Computer Science (HONS)
Faculty of Information and Communication Technology (Perak Campus), UTAR

property listing's data. This is similar to the result that has been shown in Figure 5.2.2 which displayed the total number of property in the area. Furthermore, the median price of the type of property is manually calculated by adding similar type of property and divides the number of the similar type of property. The calculated median price is similar to the result as shown in Figure 5.2.3.

CHAPTER 6: CONCLUSION

6.1 Project Review

In this project, the reason that the project was developed is to provide a more convenient ways to allow user to access property information from multiple property websites. Metasearch engine was being implemented in this project to aggregate and collect the property information from different websites. Besides that, drawing features that allow user to perform search by drawing their desire search area onto the map is included in the system. Unlike search feature found from most of the property website that only allow user to input their desired search area with name or addresses of the area. The system provides a more flexible ways and platform for user to search their desired property by using the drawing feature. Moreover, this system not only provides search capability to help user, it also provide information regarding about properties that are within the search area of user. User able to know the total number of property that is listed on the search area, the type of property and their median price within the area.

Overall, the project has successfully accomplished the objectives that are stated in the beginning.

6.2 Limitation

Although the project is accomplished and tested, there are still some limitations in this system. First, if the search area drawn by the user does not include the red markers that plotted on the map then the system unable to know what is the

location of the search area. This is due to the reason that the system needs to identify the location of search area by referring on the information of the red marker that located within the search area. These red markers hold the information about the addresses and name of the area.

Furthermore, the system only able to provide search on property data that is located on Petaling Jaya. Information about the location where the red markers plotted on the map is obtained and collected from https://maps.google.com. These information needs to be included to the system in order to make the system capable on searching property data that are located on other area.

Finally, some of the property website does not welcome bot or program to visit their webpage. In this project, propertyguru.com, one of the property website, utilize captcha image to identify whether the visitor of the website is a program or a real user. In order to extract property data and reduce the suspicious from this website, the crawler of the system is developed to crawl the page of the website with the speed that is similar to a real user. Hence, the system required time to finish extract the property data from the website. The time that needed for the system to display the search result is correspondence with the page number that needed to be crawl by the crawler of the system. If there are large number of property listing that is within the search area, the system would require a long time to display the search result.

6.3 Future Implementation

There are several enhancements that can be made to improve the proposed system. First, the proposed system only provide search feature that needed the user to draw their desired search area on the map. In order to not limit the freedom on how

user would want to perform search action. Other search feature that is commonly provided by property website should be included to the system. The user should also be able to input addresses or location of the area to perform search action.

Besides that, area information such as price trend, minimum price and maximum price of the property that are within the search area could also be provided by the system.

Finally, filters on the search result could also be included to the system. The system should provide this functionality to allow user to filter the displayed search result. User could easily view the property listing data that match their desire requirement by using this feature.

REFERENCES

Diane, T 2012. *What is a Rent Zestimate?* Available from:
<http://www.zillow.com/wikipages/What-is-a-Rent-Zestimate/>. [8 March 2015].

Folger, J n.d., *Homebuyers' walkthrough: Consideration when buying a home.*
Available from: <http://www.investopedia.com/university/homebuyers-
walkthrough/homebuyers-walkthrough3.asp >. [8 March 2015].

Investopedian.d.,*Comparative Market Analysis.*Available
from:<http://www.investopedia.com/terms/c/comparative-market-analysis.asp>. [8
March 2015].

ISTQB Exam Certification, 2015.*What is incremental model- advantages,
disadvantages and when to use it?* Available from:
<http://istqbexamcertification.com/what-is-incremental-model-advantages-
disadvantages-and-when-to-use-it/>. [9 March 2015].

JabatanPenilaiandanPerkhidmatanHarta 2011, *Valuation and property service activity.*
Available from:
<http://www.jpph.gov.my/V2/index3service.php?versi=1&no_khidmat=2&no_item=
1>. [9March 2015].

Bachelor of Computer Science (HONS)
Faculty of Information and Communication Technology (Perak Campus), UTAR

REFERENCES

Jacques, J n.d., NYC Fianance.*Cooperative/Condominium Comparable Rental Income Archives*. Available from: <http://www.nyc.gov/html/dof/html/property/property_condo_coop_comp_archive.shtml>. [8 March 2015].

Margaret, R 2011.*Heat map (heatmap)*. Avaible from: <http://searchbusinessanalytics.techtarget.com/definition/heat-map>. [8 March 2015].

Standard on Automated Valuation Models (AVMs) 2003, Available from: <http://katastar.rgz.gov.rs/masovna-procena/Files/6.Standard_on_Automated_Valuation_Models.pdf>. [8 March 2015].

Washington State Department of Licensing 2015，*State-certified real estate appraiser.* Available from: <http://www.dol.wa.gov/business/appraisers/appcertres.html>. [9March 2015].

Wickell 2015, *About Home: Home buying and selling categories.* Available from: <http://homebuying.about.com/cs/appraisals/a/real_estate.htm>. [9March 2015].

REFERENCES

Metasearch Engine, (wiki article), 11March 2015, Available from: <http://en.wikipedia.org/wiki/Metasearch_engine>. [13 March 2015].

REFERENCES