FINAL YEAR PROJECT II (TBB 4014)

PROGRESS REPORT

Online Advertising – Profitable Keyword Recommender System

by

Noor HasomyBinti Hamid

Bachelor of Technology (Hons)

Business Information System

SEPTEMBER 2013

UniversitiTeknologi PETRONAS

Bandar Sri Iskandar

31750 Tronoh

Perak DarulRidzuan

CERTIFICATION OF APPROVAL

Online Advertising – Profitable Keyword Recommender System

By

Noor Hasomy Binti Hamid

A project dissertation submitted to the

Business Information System

Universiti Teknologi PETRONAS

in partial fulfillment of the requirement for the

Bachelor of Technology (Hons)

(Business Information System)

Approved by,

_____

(MR. AHMAD IZUDDIN BIN ZAINAL ABIDIN)

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

SEPTEMBER 2013

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 have been specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

_____

NOOOR HASOMY BINTI HAMID

ABSTRACT

The main objective of this project is to create a web based system that provides a solution to find profitable keywords; Thus, increasing the efficiency of money spent in online advertisement. By using suitable algorithm and set of database, a list of profitable keyword could be produced, which inclusive the words and the suggested price to bid on. This system will help the advertisers in investing their money of the right keywords and price which increase the money spending and also the efficiency of it. To study the awareness of the online advertising among public, a set of questionnaire has been distributed among students of UniversitiTeknologi PETRONAS. The result from the questionnaire is used to study the awareness and also the importance of keywords among the public. Since the author only has less than 10 months to complete the project, she used the Rapid Application Development (RAD) as the methodology. In this report, the Gantt Chart, key milestone, along with key activities has been included as a guideline in developing the project and also, to keep track the progress of the project. Both Gantt Chart and key milestones are inclusive of both FYP 1 and FYP 2.

ACKNOWLEDGEMENT

. Praise to Allah, the most Gracious and the most Merciful'

First and foremost, my greatest gratitude to God for His blesses, given strengths and guidance to overcome challenges and difficulties during the completion of the Final Year Project.

My great appreciations to my supervisor, Mr. Ahmad Izuddin bin Zainal Abidin. Thank you for giving me chance to be one of your supervisee. The guidance and patience that I received thought the completion of the project is so much appreciated.

I would also like to express my greatest gratitude to the UTP's lecturers for the learning that I gained throughout my study years which are very handy in completing this project. Thank you for patiently teach me and never tired of answering my question pertaining to the project.

Special thanks I would like to give to Mohd Azfar bin Tommy (former GA of Internet Programming) for his guidance and patient in ensuring that I understand the use of PHP in my project. Thank You for helping me in writing the coding. Without him, I will not be able to finish it successfully and on time.

To my friends and family, thank you for your support and I appreaciate that you never give up and stop believing in me. Your support has keep me alive and help me in completing this project.

TABLE OF CONTENTS

CHAPTER 5: CONCLUSION

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND NOMENCLATURES

tf.idf	Term Frequency. Inverse Frequency
CPC	Cost Per Click
CPM	Cost Per Mille
CPV	Cost Per View
	Cost Per Visitor
GSP	Generalized Second-Price

Chapter 1:
Introduction

This Chapter discuss about the introduction and overview of the Online Advertising and the system, which include;

- Background of Study
- Significance of the Project
- Problem Statement
- Objective and Scope of study

1.1 Background of Study

Traditionally, the businesses advertise their products or services through media mass such as television, radio and other alternative, as well as printed advertisement such as banner, flyers and etc. The aim of the advertisement is basically to reach as many potential customers as possible. Many of the businesses willing to spend millions of dollar to ensure their advertisement will be appeared on air, billboard and etc.

As time flies, the emerging of the Internet as the largest global network has changed the way the businesses advertise their products or services. The transition of the Internet as platforms for military, research, government and educational, to commercial entity, have encourages the businesses to put up the web sites and used the internet to make fortunes.

Online advertising or some people might call as online advertisement; internet marketing or e-marketing is basically the marketing and promotion of products or services that are done over the Internet. In online advertisement, there are four

market players in online advertising which are advertisers, publishers, ad networks and users.

The history of online advertisement was started after AT&T bought the first ever banner ad on Wired's website in 1994. From there, the online advertisement starting to growing, and not until the year of 2004, the launch of Facebook has made a phenomenon to the world. Billions of people use Facebook as a medium of communication and to keep in touch with people all around the world. The introducing of the business page on Facebook has become a huge platform to the online businesses and the business people could see the huge potential on putting the advertisement online. With billions of people using Facebook, the possibility of being click by atleast a million of users are huge. However, it is not fair to give all credits to Facebook as Myspace, Friendster and other social Medias have been introduced way earlier than Facebook. But, the impact of Facebook on online businesses and advertisement cannot be argued. Facebook uses the pay per click system in their advertisement policy whereby the advertisers or businesses will pay on every click made by the Facebook's users. Since there are billions of people use Facebook, the chances of being clicked are huge. The businesses and advertisers are willing to invest due to this potential. Indirectly, it has been contributed to the emerging of online advertising.

In the other hand, the online advertising models also make use of the Internet's most popular applications which are email and the web. Example of popular advertising via email includes; email newsletter which is a publication, created by the business and sent out to people by request; direct email which sent to people who request an email solicitations on a particular subject; and ad-supported email which offer free email access (e.g. Hotmail) to people who would use their email readers to display

paid advertisements. Example of popular advertising via the web is banners which include static banners, animated banners, interactive banners and HTML Banners. Other than that, there are also buttons, text links, sponsorship, advertorial, interstitials and many more. Most of these advertisements can be found everywhere over the internet; on the blog, webpage, search page and many more. It is not an unusual phenomenon to see the advertisement on the internet.

There are few revenue models that have been use in online advertising. The models are chosen base on the suitability of the model with the nature of the business and also the objective and the aim of both advertisers and businesses. The revenue model that is commonly used in online advertising are CPM (Cost Per Mille), CPV (Cost Per Visitor), CPV (Cost Per View), CPC (Cost Per Click) and etc. Nowadays, it is believed that Cost Per Click is one of the famous model adopted by page owner. The advertisers will only pay for every click that the users or consumers made on the advertisement. Usually, the advertisement will direct the users or consumers to their respective page. Thus, the percentage of the users or consumers purchases their products or services are higher compare to other. Thus, the Return of Revenue (ROR) is high. Facebook is one of the examples of those who adopted Cost Per Click revenue model.

No	Revenue Model	Explanation
1.	CPM (Cost Per Mille)	Advertisers pay for exposure of their message (advertisement) to specific audience. Per Mille means per thousand impressions
2.	CPV (Cost Per Visitor	Advertisers pay for the delivery of targeted visitor to the advertisers' website.
3.	CPV (Cost Per View)	Advertisers pay for each unique user's view of an advertisement or website
4.	CPC (Cost Per Click)	Advertisers pay for each time a user clicks on their listing and redirected to their website. Only pay if the listing is clicked on.

Table 1: Revenue Model and the Explanation

Basically, there are two types of models used to determine the cost for each revenue model. There are flat-rate and bid-based. Flat-rate is quite a simple model as both advertiser and publisher agreed upon a fixed rate that will be paid for each activity (depends on revenue model). Usually, this type of advertisement is used on personal web page or blog, whereby the advertisers (usually) pay directly to the web page's or blog's owner. While the bid-based involves signing contract between advertiser and publisher; whereby it allows the advertiser to compete against other advertisers in a private auction hosted by the respective publisher. In the contract, the advertiser will inform the publisher the maximum rate that the advertiser willing to pay for a given spot, often based on a keyword.

In the context of this project, the author will actually focus on the Malay keywords by using local stop word list. The Malay keywords are chosen because most of the existence technology is focusing on the wide range of language. Thus, it has lack of sensitivity to the Malay keywords. It is believed that the system will contribute to the businesses that use Malay as their medium in communicating with their customers. As people may know, AdSense, the product of Google offer a tool that will provide a suggested bid, but the users need to key in the keywords by themselves. Other than that, in term of language selection, there is no Malay language; instead there is only Indonesian language. Unfortunately, there are many terms in Indonesian that are not used nor have different meaning in Malay language. Because of that, the effectiveness of the tool is doubt.

Last words, the growing of the online advertisement is proportion to the growing of the online business. So, as long as the buying and selling over the Internet is growing, the online advertising industries will also growing. And, to make it more promising, there is no symptom of decreasing in online businesses, in fact the

business transactions that occur on the internet has shown significant increases, especially during festive seasons.

1.2 Significance of the Project

The revenue model that most commonly used is the one that involve with bidding process. The bid-based model is commonly associated with bidding on the profitable keywords which will lead to high chances for the advertiser to appear on the web page.

This paper is proposing the 'Online Advertising – Profitable Keyword Recommender System', which is a web base system that recommend the profitable keywords that the advertisers might use in acquiring the respective keywords, especially in marketing their products or services in search engine. The system will assist the advertisers in identify which keywords they should bid on via the use of a placed text ad that will appears when people search for phrases or words related to offering. The advertisements will appear as a 'sponsored link'.

Usually, the system used by the search engine is 'pay per click', which the advertisements will appear through bidding of a series of keywords. But the advertisers will only pay the amount that the advertisers have bid for as someone clicked on the advertisements as a result of a web search.

The system will recommend the keywords that the advertisers should place the bid on by ranking the scoring of each word. The recommendation is computed based on algorithms that have been set and the inventories of the rated keywords.

Besides that, the system is actually narrowing down their scope to Malay keywords. Thus, local advertisers can trust the system more as it has higher sensitivity towards the Malay keywords, as compared to other existing systems. For information, most advertisers do not focus on specific language. In fact, they generalize it with the frequency of that respective word has been used. Since the system that will be created will focus only on Malay keywords the sensitivity and reliability of the keywords suggested is higher, compare to other.

To use this web base system, the advertisers will need to enter a short description of the products or services offering. The describtion entered must be maximum of 15 words or below. Since the system is narrowing to the Malay keyword, the advertisers are expected to enter the Malay description of their advertisement. Then, the system will generate and recommend some Malay keywords that worth to invest on base on their rank.

This will increase the chances of the advertisement to be appeared on the top as the higher the ranking, the higher the chance that people will click on the advertisement. The high ranking keyword indicates that the word is not actually rare, but also common. Thus, the system will also increase the efficiency of the money spending on the advertisement. Besides that, it can also increase the chances of investing on the most profitable keywords, which most business people are most concern of.

1.3 Problem Statement

Base on observation and study, the author realizes that the bidding process for an online advertising is mostly depends on the keywords bided by the advertiser. Other restriction might be due to tight budget, but incorrect bidding on keywords will lead to lose of the potential customers. Budget might not within the control of the advertiser (it is base on the willingness of the advertiser to spend money), but, keywords are within the control of the advertiser. However, the question is how to find profitable keywords that will maximize the income generated from online advertisement? If the advertiser bid on the right keyword, the chances of appear on the web search and get hit from the users are high. Thus, it will increase the income generated from the online advertisement.

Besides that, the efficiency of the online advertisement is also being questioned. The advertisers usually concern whether their investment on the bidding of the keyword is relevant or not. And also, whether the income generated from the advertisement is able to cover the total cost of online advertisement. Thus, it is important for the advertiser to bid on the right keyword so that, the efficiency of the investment is high.

Last and not least, the advertiser must ensure that the keyword that they have been bided is reliable as in, it is almost sure that their bid will appear as the users are searching for relevant keywords. Right computing and good summary of the products are needed, so that the objective of bidding the keywords achieved.

1.4 Objective and Scope of Study

1.4.1 Objective:

To create a web based system that provides a solution to find profitable keywords; Thus, increasing the efficiency of money spent in online advertisement

In short, the webpage will suggest the expected keywords that will most likely appear on the respective webpage or search engine for the advertiser to put a bid on it.

1.4.2 Scope of Study: **Advertiser**

There are 4 players in online advertising industry which are advertiser, publisher, users and ad networks. This project will work on the advertiser side as it will generate some keywords that the advertisers should bid on, which most likely to believe will appear as the user search for relevant keywords. So, basically the advertiser is the person announce or praise products or services in public medium, which can either be through online or offline medium such as, internet, newspaper and television; in order to influence people and public to buy or use the respective products or services. The relationships between the systems that will be created with the advertisers appear as, the result of the system is based on the input, which is the summary of the product or services that the advertiser would like to advertise on the Internet. Basically, this webpage is part of the process of advertising the product or services as it is regard as a one of the stage that the advertiser should face (which is selecting the right and appropriate keywords), before advertise their products or services.

Chapter 2:
Literature Review and/or Theory

This chapter will discuss the previous works that related to the creation of the system and the proves that algorithm used is suitable for this project.

2.1 Advertising

Advertising is a form of communication between the business (and advertiser) and the customer, with the purpose of either to attract new customers or retain the old ones. In proper definition by Bovee, advertising is a non-personal communication of information (which usually been paid), persuasive in nature about the products, services or ideas by acknowledged sponsors through various media mass (as cited in ezinearticles.com, 1992). Traditionally, the advertisement is published through media mass such as television and radio, printed advertisement such as billboard and flyers, and new media such as websites and text messages. Most commonly people advertise their products with the objectives of attracting the potential customers to try their products or spread their ideas. In short, according to Professor Dr. Philip Kotler, during his interview with Niaz Uddin for a website; etalks.me, advertising means 'finding needs and filling them profitably' (as cited in etalks.me, 2013).

The advancement in communication and technology has driven the advertising to a new way of advertisement which is through Internet which known as online advertisement; internet marketing or e-marketing. Shim (2000) stated that online advertisement is different from other medium as it enabling customers to directly interest with the advertisement. The customers can click on the advertisement for more information or even d a business transaction in the same online session. Traditionally, if the customers are interested with the product being advertise, they need to either go to the nearest shop or Google it online (base on given URL).

Problems occur when customer failed to find the respective shop or websites. By having the online advertisement, it not only counters back the problems, but also gives other advantages which are tracking the activity of the customers, targeting certain group of customers, deliverability and flexibility as Internet are available 24/77 and also interactivity between customers and the product or services offered.

2.2 History of Advertising

The diagram below is extracted from online marketing magazine which is www.marketingmag.com.au. According to the www.marketingmag.com.au (2013) via article; 'Infographic: The History of Online Advertising', the number of internet users has been increasing over the year from 16 million in 1995 to 555 million in 2002 and 1.2 billion in 2012.

The diagram also shows the starting point of online advertising which is when AT&T buy the first banner ad on Wired's website, made by Modern Media. Google had launched their Adwords in 2000 and has revolutionized ever since. According to the article also, the first search advertising keyword auction was created by goto.com in 2001 and was bought by Yahoo! on 2003. Both Google and goto.com have helping the advertising industry to revolutionize to what it is today. People of the world are referring to Adwords in assist them in advertise their products or services. As for goto.com, the idea of keyword auction has changed the way people do business in adverting. However, for the Google, they are lack of sensitivity towards other language than English and the idea of keyword auction should be used in modernize the way people advertise their products or services.

The popularity of Facebook among people of the world started when it was introduce on 2004. Billions of people uses Facebook as a way communication and it has created a huge opportunity to the businesses to go online and reach people all over the world. The ease of keep in touch and reaching people has given huge benefits to people in doing business, and with the help of post office, the shipment of products can be made. Since most of people have Facebook, the businesses started to see the opportunity of advertise their product on Facebook, either through their Facebook page or advertisement bar provided by Facebook. Since Facebook use the pay per click system, it has given he advantage to the business to invest money on advertisement online as they only pay for every click made by the consumer. From the figure, it is shown that the usage of internet is increasing especially after the social network such Facebook, Twitter had been introduced. So, as long as the social networks exist, there are always opportunity for advertisements and reaching people.

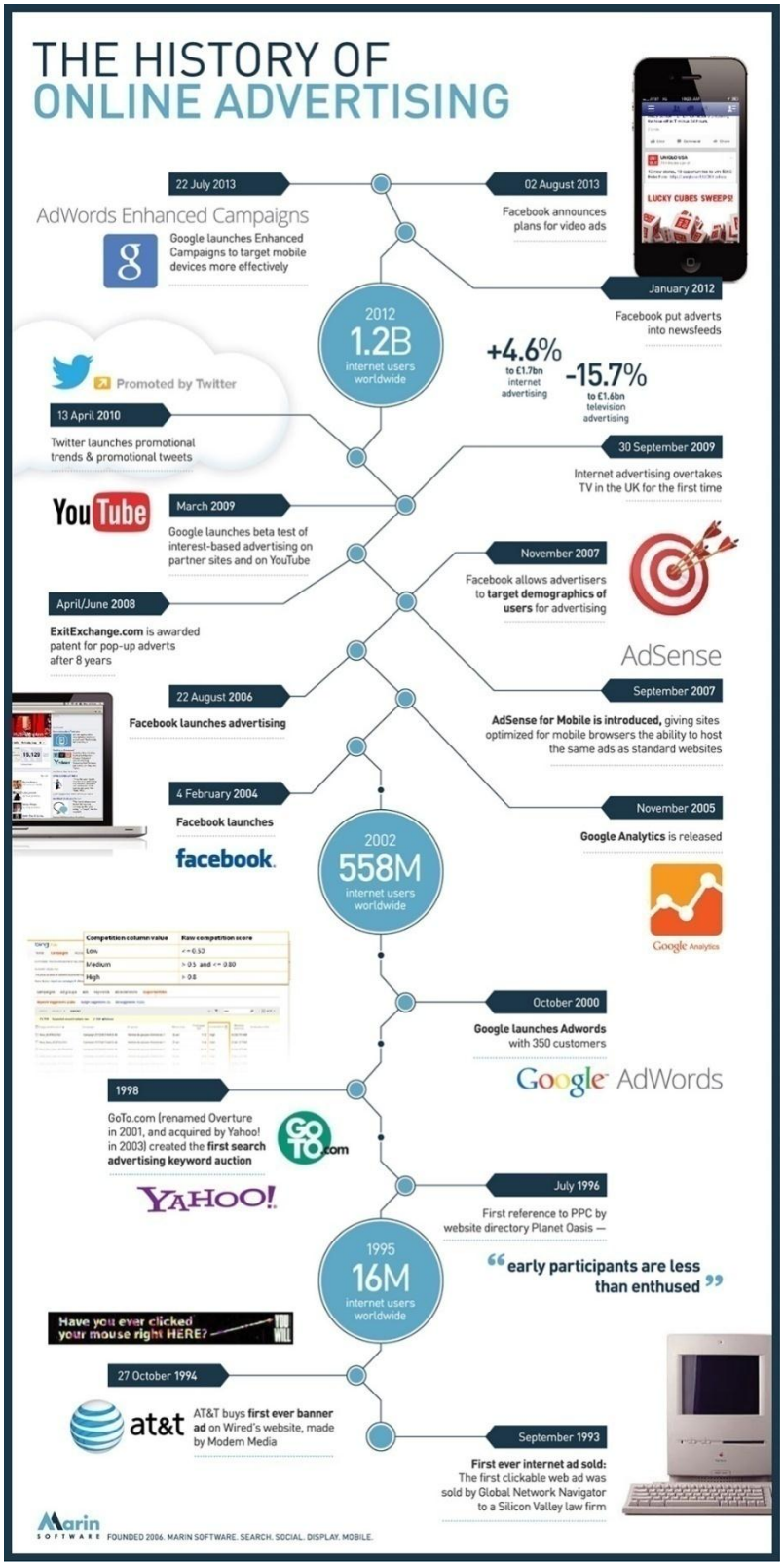


Figure 1: The History of Online Advertising

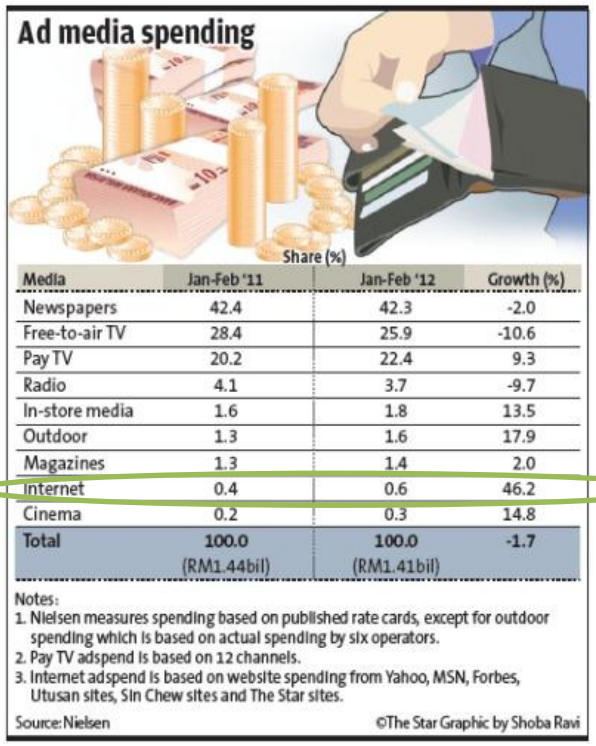
2.3 Trend in advertising

Many people believe online advertising is a growing trend, based on the average spending time on online business and devices itself. According to article that was prepared by Ho (2011) for thestar.com.my, in 2010, Malaysians had spending RM 1.8bil on online shopping. Based on last year survey, Malaysian spend more than RM100mil on travel, bill payments, entertainment and lifestyle, IT and electronics, general insurance, and fashion and beauty. Besides that, for this year, the spending on online shopping is expected to growth triple than last year. Base on this trends, and the proportion of online advertisement that growth simultaneously with the online shopping, it is believe that the online advertisement has a bright future ahead.

Figure 2: Top Seven Things Malaysia Pay for Online

As mentioned above, the online advertising is expected to have a positive growth. Base on published article written by Mahpar (2012) for thestar.com.my, the top five advertisers are Unilever, Procter & Gamble, Canon Marketing, Glaxo SmithKline and L'Oreal. In the same article, based on the data collected by Nielson Company, the author mentioned about the growth of Internet advertising. Base on the figure below, Internet Advertising shows a growth of 46.2% (base on the data collected for Jan-Feb FY2011 and Jan-Feb FY2012), which is the highest in it industry, followed by outdoor, in-store media and etc.

According to Prashant (2011), he is not surprised with the growing of Internet advertising. In fact, he quoted "We are at an inflexion point when it comes to digital spending and I see exponential growth this year and for several years to come," (as cited in thestar.com.my, 2012)



Ad media spending

Share (%)

Media	Jan-Feb '11	Jan-Feb '12	Growth (%)
Newspapers	42.4	42.3	-2.0
Free-to-air TV	28.4	25.9	-10.6
Pay TV	20.2	22.4	9.3
Radio	4.1	3.7	-9.7
In-store media	1.6	1.8	13.5
Outdoor	1.3	1.6	17.9
Magazines	1.3	1.4	2.0
Internet	0.4	0.6	46.2
Cinema	0.2	0.3	14.8
Total	100.0 (RM1.44bil)	100.0 (RM1.41bil)	-1.7

Notes:
1. Nielsen measures spending based on published rate cards, except for outdoor spending which is based on actual spending by six operators.
2. Pay TV adspend is based on 12 channels.
3. Internet adspend is based on website spending from Yahoo, MSN, Forbes, Utusan sites, Sin Chew sites and The Star sites.

Source: Nielsen ©The Star Graphic by Shoba Ravi

Figure 3: Ad Media Spending for Jan-Feb FY2011 and Jan-Feb FY2012

In the other hand, a study by mvfglobal.com (n.d.) shows that internet users from Sabah and Sarawak are more interested in travel, online games and gambling while internet user from Peninsular Malaysia tend to show great interest towards newspapers, online games and retail.

Base on this study, the marketers or advertisers can use it as a guideline to associate the content with the region. By doing this, they have a greater chances of reaching potential customers and clients.

2.4 Bidding and Keywords

The project is focus on the bidding process of the contextual ad and also finding those keywords that are believe to give revenue to the businesses or advertisers. Basically keywords are words that represent a great significance or word that act as a key to cipher or code. In the other hand, bidding is an offering of a particular price for something (dictionary.com, n.d.). Contextual ad is "a form of targeted advertising for advertisements appearing on websites or other media" (Wikipedia.com, 2007)

An article written by Williamson (n.d., p 1) gives an overview for search-based advertising services offered by search engine. As the businesses intend to advertise their products or services, they need to set a daily budget, select keywords and determines the bid price for each selected keyword. The position of the advertisement is depending on the bidding set by the business, with the ranking starting from highest to lowest. If the daily budget set is exceeding the limit, the advertisement will not be displayed.

Other mechanism of auction for online advertisement is Generalized Second- Price (GSP). According to paper written by Edelman, Ostrovsky, & Schwarz (n.d., p. 242-243), the number of ads that shown by search engine is limited. Besides that, different advertisers have different desirability on different position on the search

result page. Hence, GSP is widely used by the search engine. In GSP auction, the advertisers will choose specific keywords and put a bid on it (stating their maximum willingness to pay for a click). When a user enter a particular keyword, he or she will receive the search result along with sponsored links that shown in decreasing order of bids. Basically, the advertisement with the highest bid will be displayed on the top, second highest bid will be displayed on the second position and so on. Beautiful fact of the GPS is that, the advertiser is charged by the search engine with the amount of second highest bid.

2.5 Stop Words

Later in this paper, in Chapter 3, that is discussing the engine of the project and the flow of the engine that worked behind the system. One of the steps include, eliminating the stop words. According to the dictionary.reference.com, stop words means ‘a term you specify to exclude from your database because it occurs too frequently or because of its unimportance to the database content’. Example of the stop words are; ‘is’, ‘a’, ‘an’ and so on.

In a paper written by Abdullah, Ahmad, Mahmod and Sembok (2005) which is focus on Malay Keywords and approaches used to determine the stop words. According to them, stop words means words that frequently occurred in the documents and do not give any hint values to the content. Hence, these words will be eliminated from the set of the index terms. According to them again, there are several approaches used to determine the stop words, ranges from the manual selection to the statistically motivated methods of occurrences in order to find words with high frequently and very low frequently. But, these approaches have the same aim which is to find those that have no content values.

Salton and McGill claimed that such words (stop words) comprise around 40% to 50% of documents text words. (As cited in Abdullah et al., 2005, p. 1).

Thus, excluding these stop words will saves a huge amount of spaces in indexes and speed the process of finding the valuable keywords. Besides that, it also does not damage the retrieval effectiveness.

Table below is the lost of Malay stop words as been listed by Abdullah, Ahmad, Mahmod and Sembok (2005). The list is very useful that the author is using them for the project, for the purpose of eliminating the Malay stop words.

Ada	adakah	adakan	adalah	adanya	adapun
agak	agar	akan	aku	akulah	akupun
al	alangkah	allah	amat	antara	antaramu
antaranya	apa	apa-apa	apabila	apakah	apapun
atas	atasmu	atasnya	atau	ataukah	ataupun
bagaimana	bagaimanakah	bagi	bagimu	baginya	bahawa
bahawasanya	bahkan	bahwa	banyak	banyaknya	barangsiapa
bawah	beberapa	begitu	begitupun	belaka	belum
belumkah	berada	berapa	berikan	beriman	berkenaan
berupa	beserta	biarpun	bila	bilakah	bilamana
bisa	boleh	bukan	bukankah	bukanlah	dahulu
dalam	dalamnya	dan	dapat	dapati	dapatkah
dapatlah	dari	daripada	daripadaku	daripadamu	daripadanya
demi	demikian	demikianlah	dengan	dengannya	di
dia	dialah	didapat	didapati	dimanakah	engkau
engkaukah	engkaulah	engkaupun	hai	hampir	hampir-hampir
hanya	hanyalah	hendak	hendaklah	hingga	ia
iaitu	ialah	ianya	inginkah	ini	inikah
inilah	itu	itukah	itulah	jadi	jangan
janganlah	jika	jikalau	jua	juapun	juga
kalau	kami	kamikah	kamipun	kamu	kamukah
kamupun	katakan	ke	kecuali	kelak	kembali
kemudian	kepada	kepadaku	kepadakulah	kepadamu	kepadanya

kepadanyalah	kerana	kerananya	kesan	ketika	kini
kita	ku	kurang	lagi	lain	lalu
lamanya	langsung	lebih	maha	mahu	mahukah
mahupun	maka	malah	mana	manakah	manapun
masih	masing	masing-masing	melainkan	memang	mempunyai
mendapat	mendapati	mendapatkan	mengadakan	mengapa	mengapakah
mengenai	menjadi	menyebabkan	menyebabkannya	mereka	merekalah
merekapun	meskipun	mu	nescaya	niscaya	nya
olah	oleh	orang	pada	padahal	padamu
padanya	paling	para	pasti	patut	patutkah
per	pergilah	perkara	perkaranya	perlu	pernah
pertama	pula	pun	sahaja	saja	saling
sama	sama-sama	samakah	sambil	sampai	sana
sangat	sangatlah	saya	se	seandainya	sebab
sebagai	sebagaimana	sebanyak	sebelum	sebelummu	sebelumnya
sebenarnya	secara	sedang	sedangkan	sedikit	sedikitpun
segala	sehingga	sejak	sekalian	sekalipun	sekarang
sekitar	selain	selalu	selama	selama-lamanya	seluruh
seluruhnya	sementara	semua	semuanya	semula	senantiasa
sendiri	sentiasa	seolah	seolah-olah	seorangpun	separuh
sepatutnya	seperti	seraya	sering	serta	seseorang
sesiapa	sesuatu	sesudah	sesudahnya	sesungguhnya	sesungguhnyakah
setelah	setiap	siapa	siapakah	sini	situ
situlah	suatu	sudah	sudahkah	sungguh	sungguhpun
supaya	tadinya	tahukah	tak	tanpa	tanya
tanyakanlah	tapi	telah	tentang	tentu	terdapat
terhadap	terhadapmu	termasuk	terpaksa	tertentu	tetapi
tiada	tiadakah	tiadalah	tiap	tiap-tiap	tidak
tidakkah	tidaklah	turut	untuk	untukmu	wahai
walau	walaupun	ya	yaini	yaitu	yakni
yang					

Table 2: List of stop words

2.6 tf.idf

The algorithm used for this system is known as tf.idf (Term Frequency. Inverse Document Frequency). According to Wikipedia (2013), tf.idf is ‘a numerical statistic which reflects how vital a word is to a document in a collection’.

According to tf.idf.com (n.d) tf.idf is used to weight the importance of a word to a document in a corpus. The importance of the word increases proportionally to the number of word appears in the document. But, it is offset by the frequency of the

word in collection. Last and not least, tf.idf can be successfully used as a stop words filtering especially in text summarization and classification.

The author used this algorithm due to the stated function whereby the system could successfully function by using this method.

Chapter 3:
Methodology/Project Work

This chapter discusses the method used in developing this project, the tools used in constructing the system and other contributions that related to developing this project.

3.1 Methodology

Due to time constraint (total of less than 7 months time allocated for developing this project), the most suitable methodology to be used is the Rapid Application Development (RAD). RAD is a software development methodology that uses minimal planning in courtesy of rapid prototyping. This methodology allows the software to be written much faster and easy to change requirement. Other advantages are flexibility and adaptability to changes and the ability to reduce the project risk.

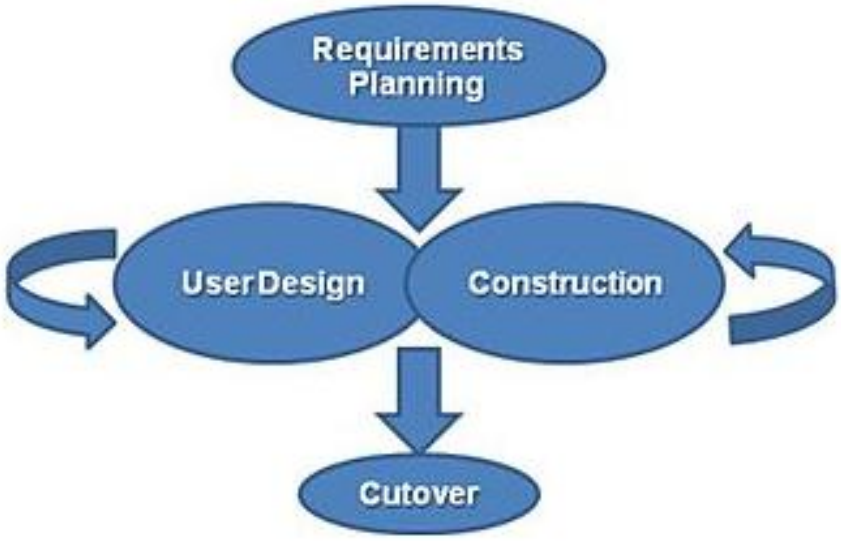


Figure 4: Phases is Rapid Application Program (RAD)

3.1.1 Requirement Planning Phase:

It is a combination of the system planning and system analysis phases of the System Development Life Cycle (SDLC). The users will discuss with the managers and IT staff, which then, agree on the business needs, project scope, constraints and system requirement. The phase ends when the team agrees on the issues and obtains management authorization to continue with the next phase.

As for this project, the developer will discuss with the supervisor regarding the needs, project scope, constraints and system requirement (base on requirement of the system and also FYP1 guidelines). The developer will continue to the next phase after both parties agrees to the key issues and have similar understanding with the project.

3.1.2 User Design Phase:

The users will communicate with the system analyst and develop models and prototypes. These models and prototypes represent all system processes, inputs and outputs. Typically, the RAD groups use a combination of Joint Application Development (JAD) techniques and CASE tools to convert users' requirements into working models. It is a continuous interactive process that allows users to understand, modify and approve the working model.

For this project, the developer will translate all the agreed needs from the previous phase into working model. The developer has a chance to continually modify the model base on latest or changing requirement.

3.1.3 Construction Phase:

This phase is similar to SDLC as it focuses on the program and application development. The users can still suggest for any modification or improvement as actual system or reports are developed. The tasks involve in this phase are programming and application development, coding, unit-integration and system testing.

As for this project, the developer will focus on the programming and application development, coding and background calculation, unit-integration, and system testing. As the phase goes on, the developer can still do modification or improvement to the system or reports.

3.1.4 Cutover Phase:

This final phase is similar to the implementation phase in SDLC which also include data conversion, testing, changeover to the new system and user training. The entire process will be compress and consequently, the new system will be in operation much sooner.

For this project, the tasks involved are data conversion, full-scale testing, system and system changeover.

3.2 Tools

Below are the tools that will be use in developing the project.

3.2.1 AdWords

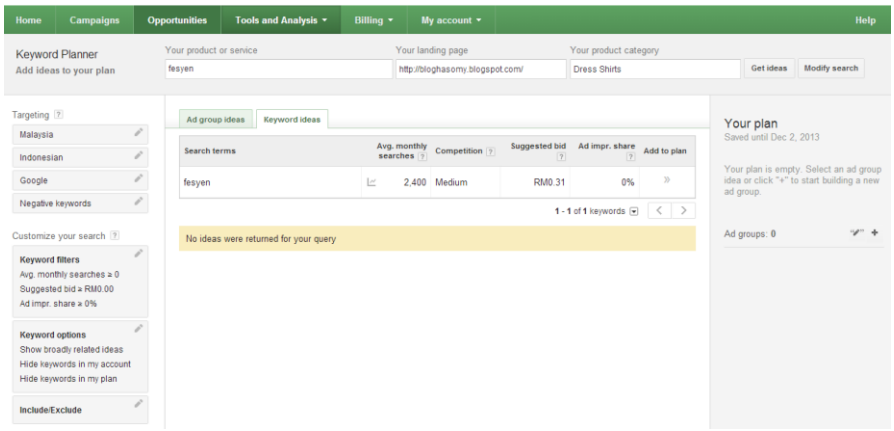


Figure 5: Keyword Planner in AdWords

AdWords is used to study any lack in the system which the author could use for her FYP. Base on observation, there is no sign of keyword recommender system. Thus, the author decided to continue with project. Basically AdWords is a campaign management which specialized in online advertising. They also provide a tool to help the advertisers to select the keywords that they want to use with its' budget. This tool is known as 'Keyword Planner".

3.2.1.1 Keyword Planner

Keyword Planner is a tool provided in the AdWords for users to key in their price and get;

- Average monthly searches
- Level of competition on the keyword
- Suggested bid

3.2.2 Wampserver

Figure 6: Wampserver's Logo

WampServer is a Windows web development environment which allows users to create web applications with Apache2, PHP and MySQL database. Besides, PhpMyAdmin allows the users to manage the database easily. Functionalities provided by wampserver are divided into two;

3.2.2.1 Left click on wampserver's icon; the users will be able to;

- Manage the users' Apache and MySQL services
- Switch online or offline; either give access to everyone or only localhost.
- Install and switch Apache, MySQL and PHP releases
- Manage users' servers settings
- Access users' logs
- Access users' settings files
- Create alias

3.2.2.2 Right click on wampserver's icon; the users will be able to;

- Change WampServer's menu language
- Access wampserver's page

3.2.3 Notepad

Figure 7: Notepad's Logo

Notepad is a common text-only or plain text editor. The resulting files, saved in the .txt extension; have no format tags or styles, which makes the program suitable for editing system files that are to be used in a Disk Operating System environment and source code for later compilation or execution, usually through a command prompt.

For this project, the author used this software to write the system in PHP language.

3.2.4 PHP

Figure 8: PHP's Logo

PHP is a server-side scripting language, designed for web development and also used as general purpose programming language. It was created

by Rasmus Lerdorf in 1995. Originally, PHP is stands for *Personal Home Page*, but now, it is known as *Hypertext Preprocessor*. PHP code is interpreted by a web server that has a PHP processor module that will be resulting into a web page.

The author learnt the PHP language previously and has decided to use it in creating the web site. As the project is focusing on proving the system and algorithm, the web site developed will be much simple.

3.3 System Architecture

System architecture and storytelling way of explaining the architecture of the system could help to clarify how things are currently working and how they could be improved on. It also helps in clarifying the key elements and most importantly, it gives a clear picture of the process of the system. Figure below illustrate the basic overview on how the system works.

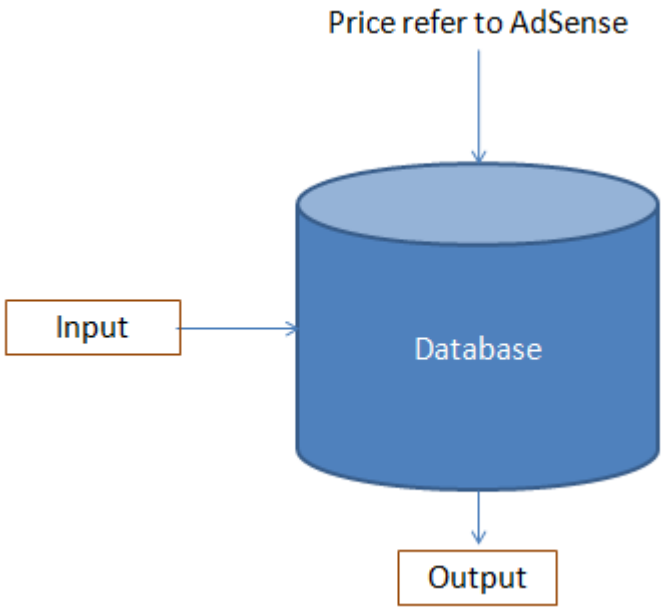


Figure 9: System Architecture

Based on the diagram above, the system basically has three main elements that will serve the system. There are the database, the input and also the output.

The input is basically the summary of the advertisement and also the maximum budget per click, provided by the advertisers. Currently, the suggested number of the summary is 15 words. The longer the query, the longer the time that the system will took to process.

In the database system, there will be the list of stop words which will eliminate those word that is presence in the database. Thus, irrelevance or word with no value will be successfully deleted from the candidate list.

The input and output is illustrated in the figure below.

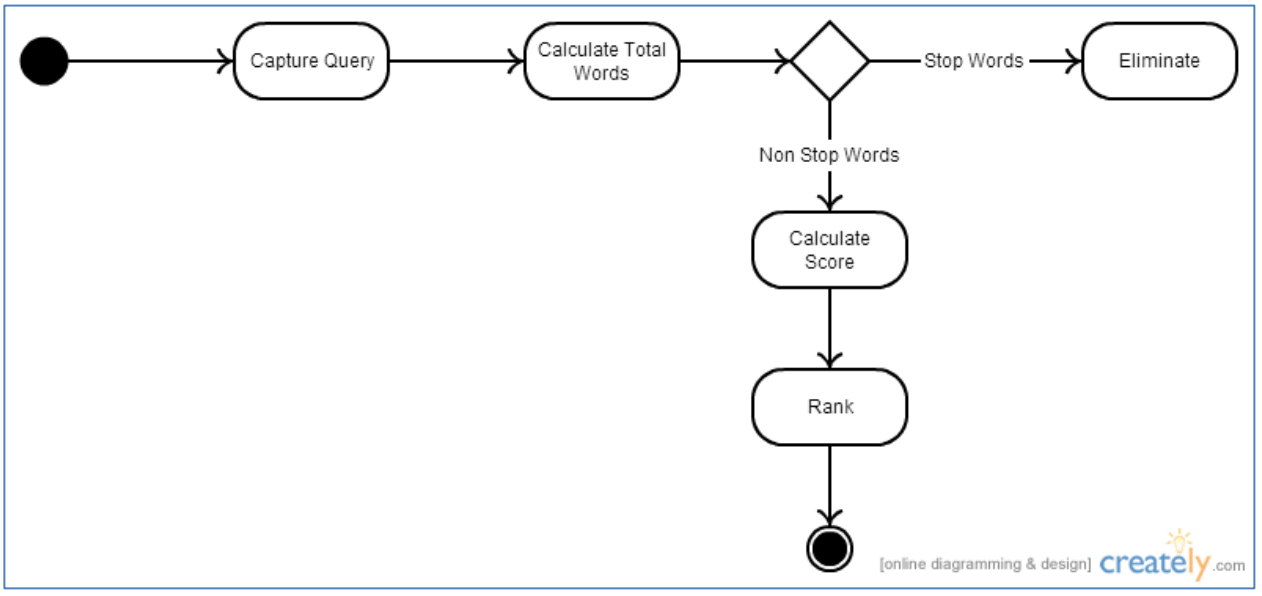


Figure 10: Process Flow

From the above figure, there are 6 steps involved in this system and below are the explanations, mathematical formula used, and also the source code.

For the input, the users are required to key in the query which briefly explains their products or list of suggested keywords that they would like to purchase. The suggested numbers of keywords that they should key in is, maximum of 10-15 numbers. This is because, the longer the words, the longer the time that the system required to process the data.

Below is the source code for the query.

```
<html>
    <head>
    </head>
    <body>
        <form name="form_query" method="get" action="tfidf.php">
            <input type="text" name="query" size="100" />
            <input type="submit" value="Submit" />
        </form>
    </body>
</html>
```

Table 3: Source Code for Query

Then, the system will calculate the total words keyed in.

```
$word = $_GET["query"];

$a = explode(" ", $word);
$d = array();
$count = 0;
$g_search_result_array = array();
$val_n = 0;
$tf = 0;
$idf = 0;
$tfidf = 0;
$rank = array();
```

```
$ranking = 1;

for($i = 0; $i < count($a); $i++) {
	$a[$i] = preg_replace("/[^A-Za-z0-9\-]/", "", $a[$i]);
	$a[$i] = strtolower($a[$i]);
}

for($i = 0; $i < count($a); $i++) {
	echo $a[$i];
	echo "<br />";
}

echo "Total words: ".count($a)."<br />";

echo "<br />";
```

Table 4: Source Code for Total Words

After that, the system will eliminate the all the stop words base on the list of stop words provided.

```
for($i = 0; $i < count($a); $i++) {
	for($j = 0; $j < count($stop_word); $j++) {
		if($a[$i] == $stop_word[$j]) {
			$count++;
		}
	}
	if($count == 0) {
		echo $a[$i]." is choosen word";
		echo "<br />";

		array_push($d, $a[$i]);
	}
	else if($count > 0) {
		echo $a[$i]." is stop word";
		echo "<br />";
	}
	$count = 0;
}
```

Table 5: Source Code for Stop Words

And the source code for storing the list of stop words is stored in the file name **'include'** and the document is named as **'stop_word.php'**

After identify the stop words, a list of candidates (keywords) is produced.

```
echo "<br />The choosen word: ";
        for($i = 0; $i < count($d); $i++) {
                echo $d[$i]."\n";
}
```

Table 6: Source Code for Candidates

Then, the score of the candidate is calculated. The idea is, to use tf.idf in finding the profitable keywords. The tf.idf value is actually, escalates proportionally to the number of times a word appears in the corpus. But, it is offset by the frequency of the word in the corpus, which helps to indicate that some words are more common than others. In calculating the score, the author used the google.com as the basis in getting the number of document found for each query or input. Below is the calculation involved in calculating the score of each candidate.

tf = total number of query/ input keyed in

df = no of document found in google.com

N = the highest value of the df (compare with all candidate)

idf = $log \log_{10} \frac{N}{df}$

tf.idf = tf ×idf

Table 7: Calculation for tf.idf

And the source code is;

```
$vals = array_count_values($d);

        foreach($vals as $t => $no_of_t) {
                //echo "No <b>".$t."</b> occur in d: ".$no_of_t."<br />";
                $g = getGoogleSearchResult($t);
                $g = preg_replace("/[^0-9]/", "", $g);
                //echo $g."<br />";
                array_push($g_search_result_array, $g);
        }
```

```
$val_n = max($g_search_result_array);
echo "N value: ".$val_n;
echo "<br /><br />";

foreach($vals as $t => $no_of_t) {
	echo "No. of <b>".$t."</b> occur in d: ".$no_of_t."<br />";

	//tf
	$tf = 1+log10($no_of_t);
	echo "tf : ".$tf;
	echo "<br />";

	//google search result
	$g_search_result = getGoogleSearchResult($t);
	$g_search_result = preg_replace("/[^0-9]/", "",
$g_search_result);

	//idf: here using google api but not available. this approach
may violate google tos
	$idf = log10($val_n/$g_search_result);
	echo "idf : ".$idf." (google.com.my search result (df) :
".$g_search_result.")";
	echo "<br />";

	//tfidf
	$tfidf = $tf * $idf;
	echo "tfidf : ".$tfidf;
	echo "<br />";

	$rank[$t] = $tfidf;

	echo "<br />";
}
```

Table 8: Source Code for Scoring

To call the function used to get the number of document found in google.com, the author keep the document named **'google_search_result_func.php'** in the file named **'include'**. The source code is;

```
<?php
ini_set('max_execution_time', 300);

function getGoogleSearchResult($str) {
        $content = file_get_contents('http://www.google.com.my/search?q='.$str);
        preg_match('/About (.*) results/i', $content, $matches);
        return !empty($matches[1]) ? $matches[1] : 0;
}
?>
```

Table 9: source code for scoring II

Lastly, is the ranking of the candidates. The highest ranking indicates that the word is rare and common (high tendency to be searched). If the word is common but not rare, it will not be in the highest position, same goes to the words that are too rare but not common. The higher the score (higher ranking), the higher the tendency of people to search the words.

```
arsort($rank);
        foreach($rank as $t => $tfidf) {
                echo $ranking.". ".$t." ".$tfidf."<br />";
                $ranking++;
        }
```

Table 10: Source Code for Rank

3.4 Gantt Chart and Key Milestones

Below is the Gantt Chart for the FYP 1 which has been successfully completed during FYP 1. For the phase 2 of the FYP, the author will focus on the table 7 which is Gantt Chart for FYP 2.

Activity/ Week	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Understand the jargons and methodology														
Literature review and research on relevant subject														
Study the background calculations (computational calculation)														
Develop ads system flowchart, database model, database design														
Design user interface (website) and create content for website														
Prepare Gantt chart for task accomplishment														
Develop key milestones														

Table 11: Gantt Chart for FYP 1

Table 7 is the Gantt Chart that was prepared early on during FYP 1 for FYP 2 purposes. Since the author is yet to have a meeting with the supervisor, some amendment is believed will be make after the meeting.

Activity/ Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Study the background calculations/ algoritm (computational calculation)														
Create the algorithm														
Improve ads system flowchart, database model, database design														
Design user interface (website) and create content for website														
Improvement and Tuning														
Meeting with supervisor														

Table 12: Gantt Chart for FYP 2

Below is the table for Key Milestone of FYP 1. The table is constructed base on the guidelines provided by the FYP 1 committee. All guidelines given has been successfully followed and completed during FYP 1 and for the FYP 2, the author will put emphasis on the table 9.

Activity/ Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Title selection/ proposal														
Submit proposal to research cluster														
Extended Proposal														
Viva: Proposal defense and Progress Evaluation														
Interim Report														
Meeting with supervisor														

Table 13: Key Milestones for FYP 1

Below is the table for Key Milestone of FYP 2. The table is constructed base on the guidelines provided by the FYP 2 committee.

Activity/ Week	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Submission of Progress Report														
Pre Sadex														
Dissertation														
Sadex														
VIVA														

Table 14: Key Milestones for FYP 2

3.5 Project Activities

The project activities are based on the methodology used, Rapid Application Development.

3.5.1 Requirement Planning

3.5.1.1 Understanding the jargons

3.5.1.2 Literature review and research on relevant subject

3.5.1.3 Prepare survey question

3.5.1.4 Prepare Gantt chart and project activities

3.5.2 User Design

3.5.2.1 Design website

3.5.2.2 Study the background calculations (computational calculation)

3.5.2.3 Develop ads system flowchart, database model, database design (use synhetic data)

3.5.2.4 Prove the calculation is reliable to be used

3.5.3 Construction

3.5.3.1 Develop prototype

3.5.3.2 Evaluate system functionality

3.5.4 Cutover

3.5.4.1 Prototype is ready

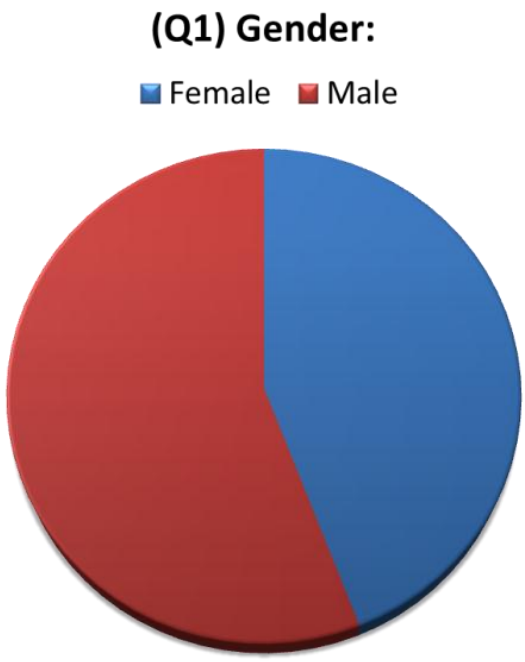
3.5.4.2 Recommendation for future study

Chapter 4:
Result and Discussion

This chapter discusses about the survey that was conducted during Final Year Project 1 and the results from the survey, and project prototype that has been developed.

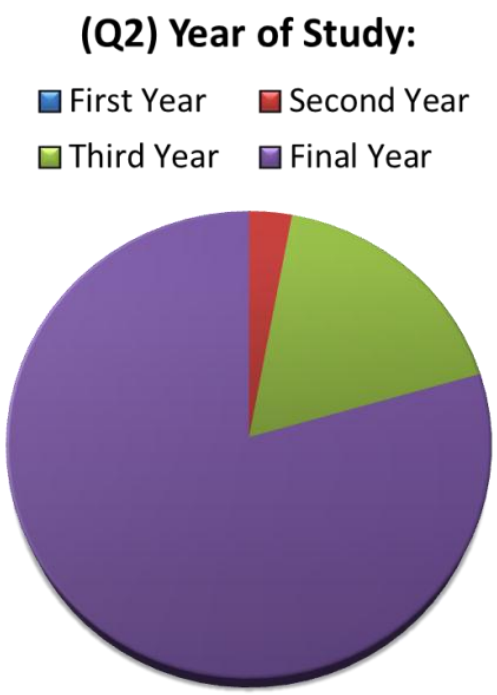
4.1 Survey Analysis

A survey has been conducted among students of UniversitiTeknologi PETRONAS which covers all courses and various years of studies. The main objective of this survey is to study their awareness on online advertising. The survey was conducted online; by spreading the emails to student of Universiti Teknologi PETRONAS.

<table>
<tr><td>No</td><td>Analysis</td></tr>
<tr><td>1.</td><td>


<table>
<tr><td>Female</td><td>15</td><td>44%</td></tr>
<tr><td>Male</td><td>20</td><td>56%</td></tr>
</table>
Analysis:
Question 1 is referring to the gender of the respondents. Base on the result of the survey, 56% is male and another 44% is female. Hence, the majority of the respondents are male.
</td></tr>
</table>

2.

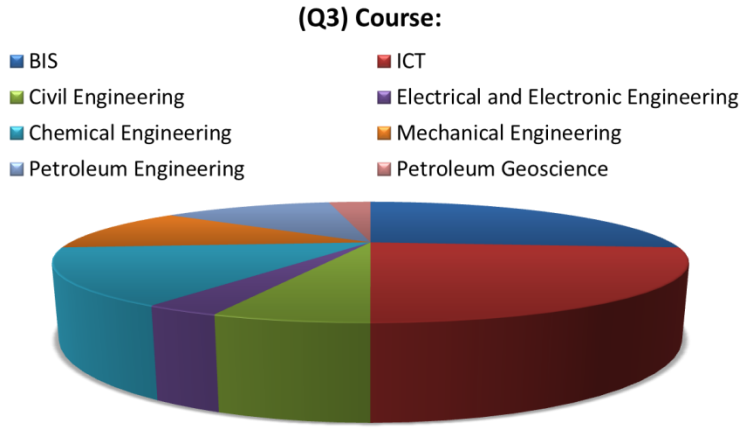


First Year	**0**	0%
Second Year	**1**	3%
Third Year	**6**	18%
Final Year	**28**	79%

Analysis:

Since the targeted respondents are all students, question 2 is referring to the year of study of the respondents. Unfortunately, there is no first year student, 3% of the respondents are second year, 18% are on their third year and 79% are the final year students. Hence, the majority of the respondents are the final year students.

3.

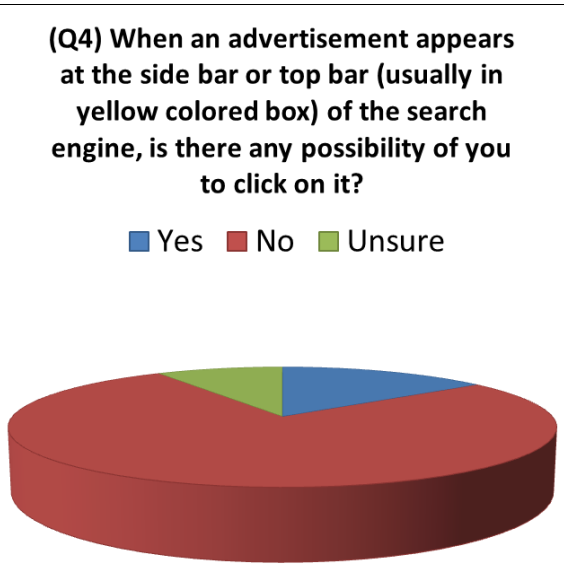


BIS	**9**	26%	Chemical Engineering	**5**	15%
ICT	**8**	24%	Mechanical Engineering	**4**	12%
Civil Engineering	**2**	6%	Petroleum Engineering	**4**	12%
Electrical and Electronic Engineering	**1**	3%	Petroleum Geoscience	**1**	3 %

Analysis:

Question 3 asked about the respondents' course and field of study. Basically, the survey has reached students of all courses. The top 3 of the courses are BIS with 26%, ICT with 24% and Chemical Engineering with 15%.

4.



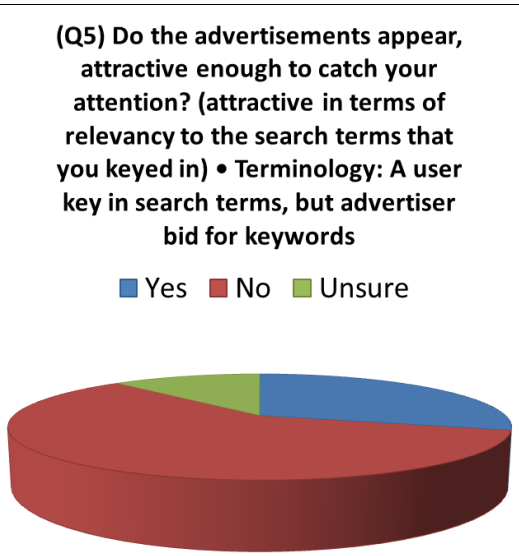
Yes	**5**	15%
No	**27**	76%
Unsure	**3**	9%

Analysis:

Question 4 asked about the willingness and possibility of the respondents to click on the contextual ad. Majority of the respondents (76%) answered 'No', 15% answered 'Yes', while another 9% answered 'Unsure'.

HYPOTHESIS: The answer of 'yes' and 'no' will be equal as people will click the advertisement base on the content they provided.

5.



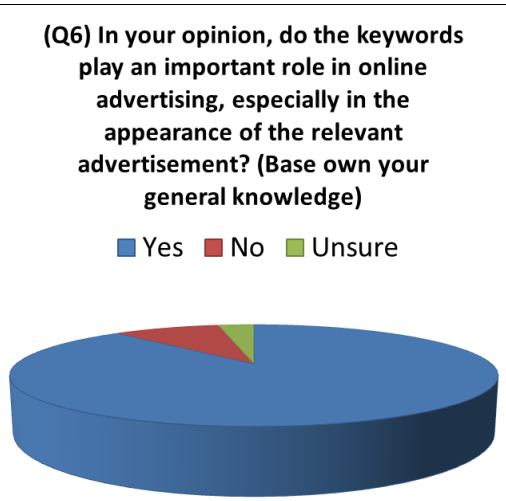
Yes	**10**	29%
No	**21**	59%
Unsure	**4**	12%

Analysis:

Question 5 is referring to the attractiveness of the advertisement and whether they are able to attract the respondents/ users to click on it, or not. Base on the analysis, 59% answered 'No', 29% answered 'Yes' while the rest answered 'Unsure'.

HYPOTHESIS: The answer of 'yes' and 'no' will be equal as people will click the advertisement base on the content they provided.

6.



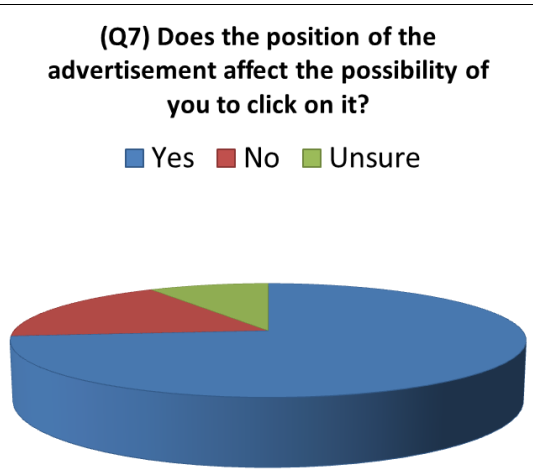
Yes	**31**	88%
No	**3**	9%
Unsure	**1**	3%

Analysis:

Question 6 brings up the importance of the keywords in online advertising. Base on the survey, majority of the respondents (88%) answered 'Yes', only 9% answered 'No', while the rest (3%) answered 'Unsure'.

HYPOTHESIS: The answer of 'yes' and 'no' will be equal as the summary of the advertisement play an important role in influencing people to click on it.

7.



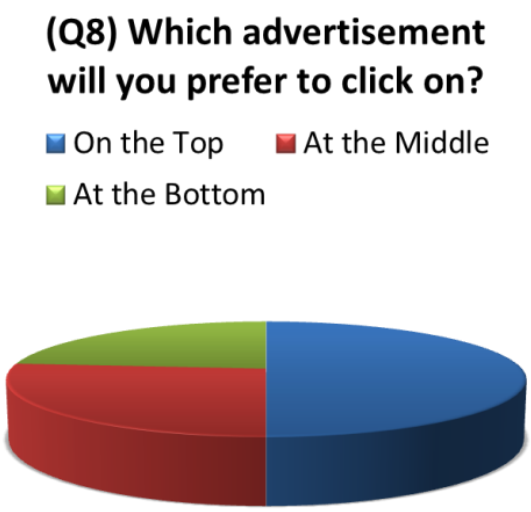
Yes	**26**	74%
No	**6**	18%
Unsure	**3**	9%

Analysis:

Question 7 referring to the position of the advertisement and the possibility of the advertisements to be clicked on. Base on the survey, majority of the respondents (74%) answered 'Yes', only 18% answered 'No', while the rest (9%) answered 'Unsure'.

HYPOTHESIS: The answer of 'yes' and 'no' will be equal as the position of the advertisement will not affect the possibility of being click on, but people will click the advertisement base on the content provided.

8.



On the Top	**18**	50%
At the Middle	**9**	26%
At the Bottom	**8**	24%

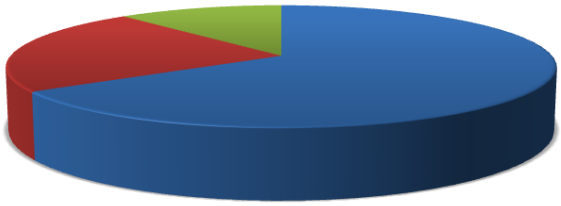
Analysis:

Question 8 brings up the positioning of the advertisements. From the survey, 50% answered 'On the Top', 26% answered 'At the middle' and 24% answered 'At the Bottom'.

HYPOTHESIS: The answer of 'on the top', 'at the middle' and 'at the bottom' will be equal as the position of the advertisement will not affect the possibility of being click on, but people will click the advertisement base on the content provided.

9.

(Q9) If the advertisement appears is relevance to what you're looking for, will you click on it?

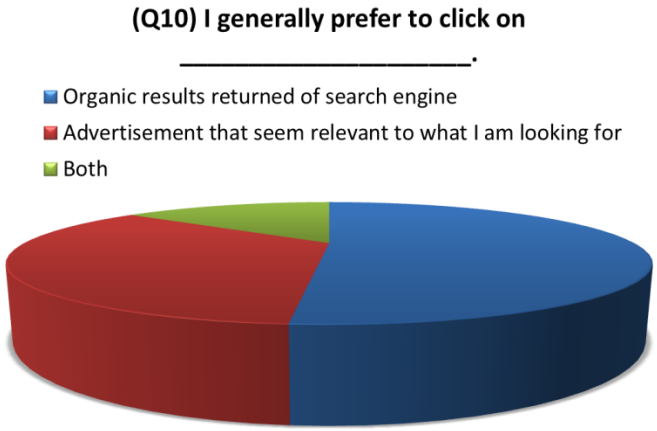


Yes	**23**	66%
No	**8**	23%
Unsure	**4**	11%

Analysis:

Question 9 asked the possibility of the advertisement to be clicked on if the advertisement is relevance to what are you looking for. The result of the survey stated that, most of the respondents (66%) answered 'Yes', only 23% answered 'No', while the rest (11%) answered 'Unsure'.

10.



Organic results returned of search engine	**18**	51%
Advertisement that seem relevant to what I am looking for	**12**	34%
Both	**5**	14%

Analysis:

Question 10 asked the preference of the respondents, whether they prefer to click on the organic result of the search engine, the advertisement or both. Base on the survey, 51% of the respondent answered 'Organic results returned of search engine', 34% answered 'Advertisement that seems relevant to what I am looking for', while another 14% answered 'Both'.

HYPOTHESIS: The answer of 'yes' and 'no' will be equal as people will click the advertisement if the summary or content of the advertisement is relevant to what they are looking for.

Table 15: Survey Analysis

4.2 Project Prototype

To stimulate the idea of the system, the author proposed a web based system. This web base system will act as an agent that will assist the advertisers in finding profitable keywords.

The author will used the PHP scripting language, notepad as tool to write and also wampserver as the server to create a dynamic website. The websites is in a very simple form as the purpose of creating the website is more on proving the algorithms and system proposed. The more complex and dynamic website will be later proposed on the Recommendation section.

This system, as mention previously, consists of six steps which are;

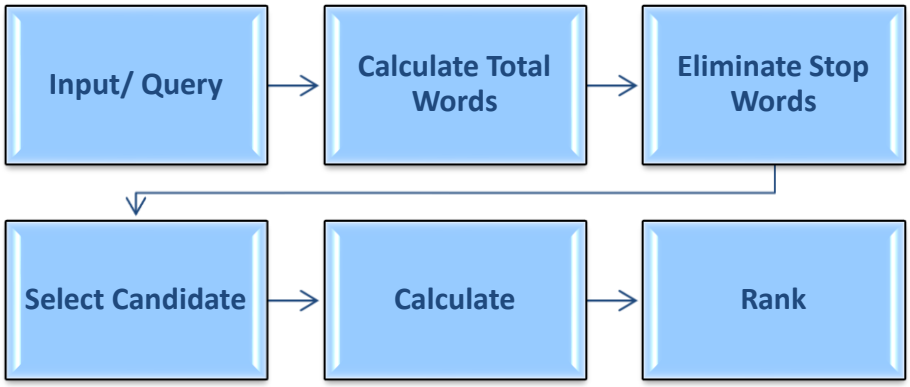


Figure 11: Six Steps of Keyword Recommender System

Figure 12: The Query

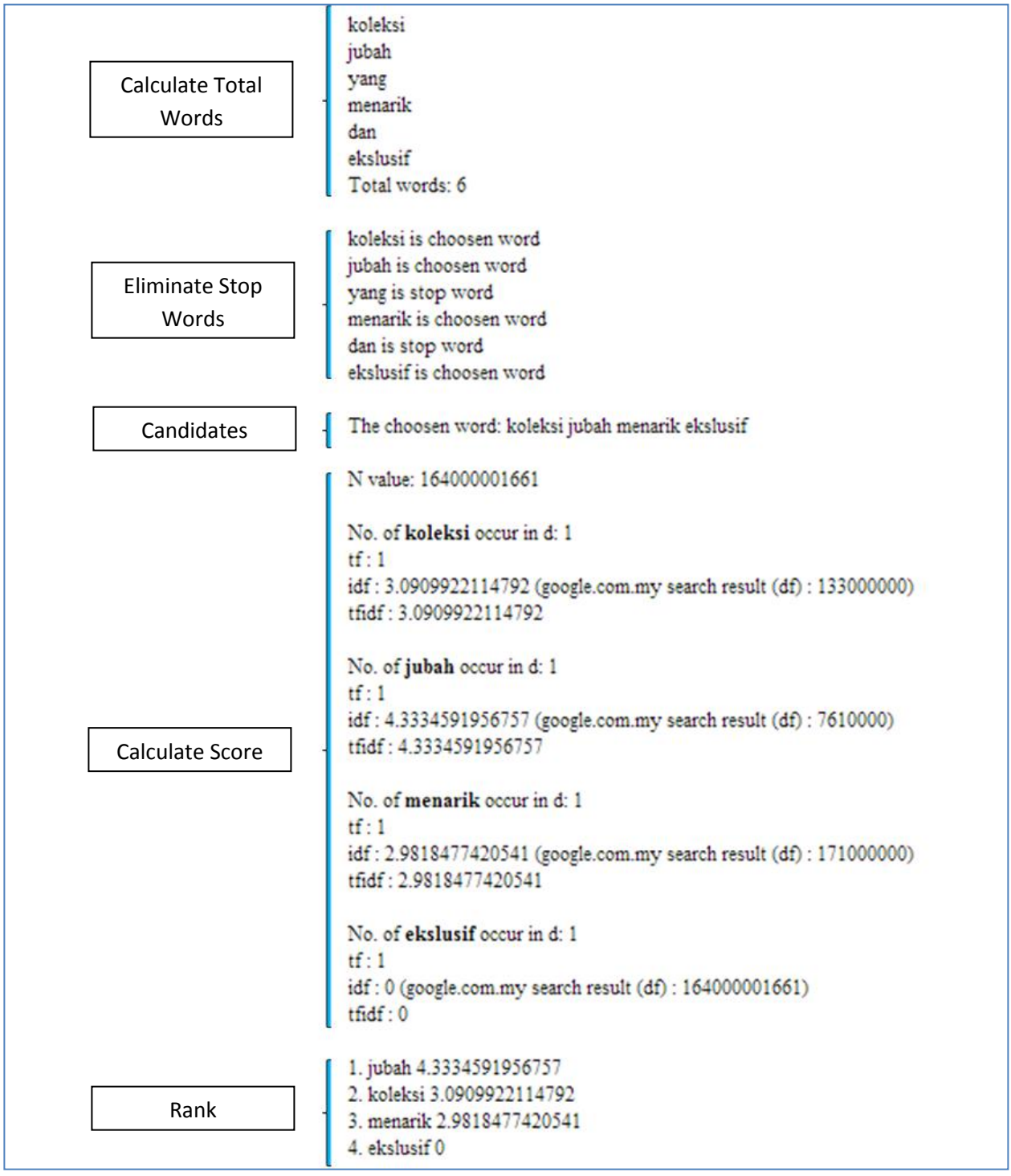


Figure 13: The Result

First, the user will be required to enter the query. The query is limited to 10-15 words as in online adverting, fewer words are used and there's no need of long explanation. Therefore, a very short and precise summary of the products sold are used. For example, the user has keyed in 'koleksi jubah yang menarik dan eksklusif'

The system will calculate the total words, which in case in 6.Then using the Stop Word method, the system will eliminate those irrelevance words that have no value. Pronounce words and conjunction words are included as irrelevance words. To be more precise, the list of stop words (as mentioned in Literature Review) is used. Those words will be eliminated precisely. For this case, the word 'yang' and 'dan' have been eliminated as there bring no value to the sentence.

The main algorithm used in the system is tf.idf (Term Frequency. Inverse Frequency). The purpose of using tf.idf is to select the words with high frequency. By using tf.idf, the words that have been not be eliminated (candidates) is scored. (refer above table)

Lastly, the keywords will be ranked base on their score. The higher the score, the higher the rank of the keywords. The highest position indicate the words is common (high query) and also rare. Base above query, the word 'jubah' has the highest score, followed by 'koleksi' 'menarik' and 'eksklusif'.

Chapter 5:
Conclusion

This Chapter wether the project has achieved it's objective or not and also future work that need to be done.

5.1 Relevancy to the Objective

Relevancy of the system to the objective of the system is based on the outcome of the system. The objective of the system (as mentioned in Chapter 1) is 'to create a web based system that provides a solution to find profitable keywords; Thus, increasing the efficiency of money spent in online advertisement. 'The outcome or result that will be generated is the keywords that is believed to have high chances to be click on, if the advertisers invested on them. In the other word, the return on revenue will be high. High return on revenue indicates that the expenses (on investing on right keywords) have been managed efficiently. Most businesses are concern with their return on revenue every time their investing their money on something. Simple words, by using the system, the businesses could maximize their return on revenue.

Besides that, the main focus in developing the system and creating the website is to find Malay profitable keywords. Hence, the system is believed to relevance to the objective of the project. Besides that, the reason the website and system to narrow down their scope into Malay keywords is because most of the existing system that available have wide scope thus there are lack of the sensitivity towards the Malay words. Besides that, by narrowing down the scope, the website will have competitive advantage in competing with the existing website and also to have a sense of locality in the system developed.

5.2 Suggested Future Work

For future work and recommendation, the author would like to suggest the continuation of this system as the algorithm has been proven as correct and reliable. The system should be more complex by adding some features that will attract people to visit the website. Since the author is focusing on proving the algorithm, the website features are weighted on the practicality only.

Besides that, the author would also like to suggest the use of own data set, instead of using Google's. For now, the author is using Google's dataset due to time constraints and as she only focuses on proving the algorithm.

As mention earlier, the algorithm has been correct, thus the system should be really make up and use by advertisers in finding profitable keywords.

References

Abdullah, M, T., Ahamd, F., Mahmod, R., &Sembok, T, M,T. (n.d). Improvement of Malay Information Retrieval Using Local Stop Words, 1-7.

Bourdon, R (n.d) Wampserver. Retrieved November 20, 2013, from http://www.wampserver.com/en/

dictionary.com (2011) Advertiser. Retrieved October 11, 2013, from http://dictionary.reference.com/browse/advertiser

Dilhan, S. (2011) Advantages and Disadvantages of RAD (Rapid Application Development). Retrieved June 27, 2013, from http://sameeradilhan.com/advantages-and-disadvantages-of-rad-rapid-application-development

Edelman, B., Ostrovsky, M., &Schwarz, M. (n.d). Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords, 242-243.

entrepreneur.com (2011) Online Advertising. Retrieved June 27, 2013, from http://www.entrepreneur.com/encyclopedia/online-advertising 24

Global Customer Acquisition (2011) Regional Internet Trends. Retrieved June 27, 2013, from http://www.mvfglobal.com/malaysia

Ho, S. (2011) Malaysians spent RM1.8bil shopping online in 2010. Retrieved June 27, 2013, from http://thestar.com.my/news/story.asp?file=/2011/4/22/nation/8534221&sec=nation

investopedia.com. (n.d) Return on Revenue - ROR. Retrieved October12, 2013, from http://www.investopedia.com/terms/r/returnonrevenue.asp

Jain, R. (n.d) Definition. Retrieved June 27, 2013, from http://ezinearticles.com/?Advertising-As-A-Tool-Of-Communication&id=1228665

Mahpar, MH. (2012) Malaysia's adex gathers momentum in February. Retrieved June 27, 2013, from http://thestar.com.my/news/story.asp?file=/2012/3/19/business/10938964&sec=

Marketing. (2013) Infographic: The History of Online Advertising. Retrieved October 12, 2013, from http://www.marketingmag.com.au/news/infographic-the-history-of-online-advertising-45065/#.UlfHdlBpnGE

Mitchell, B. (n.d) Apache. Retrieved June 27, 2013, from http://compnetworking.about.com/cs/webservers/g/bldef_apache.htm

opensitesearch (2007) Stopwords Definition. Retrieved August 19, 2013, from http://opensitesearch.sourceforge.net/docs/helpzone/dbb/dbb_50-10-12r.html

openx.com. (2011) FAQs: General Questions. Retrieved June 27, 2013, from http://www.openx.com/support/faqs/general

Rusmevichientong, P., &Williamson, D.P. (n.d). An Adaptive Algorithm for Selecting Profitable Keywords, 1, 1-2.

Shim, SW. (2000) Definition. Retrieved June 25, 2013, from http://iml.jou.ufl.edu/projects/fall2000/shim/defi.htm

tf.idf.com (n.d) What does tf-idf mean?. Retrieved November 30, 2013, from http://www.tfidf.com/

Uddin, N. (n.d) Finding needs and filling them profitably. Retrieved June 27, 2013, from http://etalks.me/philip-kotler-marketing-for-better-world/

wikipedia (2011) Contextual Advertising. Retrieved June 27, 2013, from http://en.wikipedia.org/wiki/Contextual_advertising

wikipedia (2013) Notepad (software). Retrieved November 22, 2013, from http://en.wikipedia.org/wiki/Notepad_(software)

wikipedia (2013) PHP. Retrieved November 22, 2013, from http://en.wikipedia.org/wiki/PHP

wikipedia (2013) WordPress. Retrieved August 21, 2013, from http://en.wikipedia.org/wiki/ WordPress

wikipedia (2013) tf-idf. Retrieved August 21, 2013, from http://en.wikipedia.org/wiki/ Tf-idf