ECOMMERCE USING SEARCH ENGINE OPTIMIZATION (ESEO)

YEAP JIA WEI

BACHELOR OF COMPUTER SCIENCE
(SOFTWARE ENGINEERING)
UNIVERSITI MALAYSIA PAHANG

UNIVERSITI MALAYSIA PAHANG

BORANG PENGESAHAN STATUS TESIS

JUDUL: Ecommerce Using Search Engine Optimization

SESI PENGAJIAN: **2011 / 2012**

Saya: YEAP JIA WEI
(HURUF BESAR)

mengaku membenarkan tesis (Projek Sarjana Muda/~~Sarjana/Doktor Falsafah~~)* ini disimpan di Perpustakaan Universiti Malaysia Pahang dengan syarat-syarat kegunaan seperti berikut:

1. Tesis adalah hakmilik Universiti Malaysia Pahang
2. Perpustakaan Universiti Malaysia Pahang dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. **Sila tandakan (✓)

☐	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
☐	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
☑	TIDAK TERHAD	

Disahkan oleh

(TANDATANGAN PENULIS)

Alamat Tetap:
B-5-5 Awana Puri, Jalan 15/119,
56000 Cheras, W.P. Kuala Lumpur.

Tarikh : 13th June 2012

(TANDATANGAN PENYELIA)

Nama Penyelia:
Dr. Adzhar bin Kamaludin

Tarikh : 13th June 2012

CATATAN:
* Potong yang tidak berkenaan.
** Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh tesis ini perlu dikelaskan sebagai SULIT atau TERHAD.
*** Tesis dimaksudkan sebagai tesis bagi Ijazah Doktor Falsafah dan Sarjana secara penyelidikan, atau disertasi bagi pengajian secara kerja kursus dan penyelidikan, atau Laporan Projek Saujana Muda (PSM).

ECOMMERCE USING SEARCH ENGINE OPTIMIZATION

YEAP JIA WEI

A report submitted in partial fulfillment of the requirements
for the award of the degree of
Bachelor of Computer Science (Software Engineering)

Faculty of Computer Systems & Software Engineering
UNIVERSITY MALAYSIA PAHANG

JUNE, 2012

STUDENT'S DECLARATION

"I hereby declare that the work in this project is my own except for quotations and summaries which have been duly acknowledged. This project has not been accepted for any degree and is not concurrently submitted for award of other degree."

Signature : …….....
Name : Yeap Jia Wei
ID Number : CB09132
Date : June 13th, 2012.

SUPERVISOR'S DECLARATION

"I hereby declare that I have checked this project and in my opinion, this project is adequate in terms of scope and quality for the award of the degree of Bachelor of Computer Science (Software Engineering)."

Signature : …….....
Supervisor : Dr. Adzhar bin Kamaludin
Position : Deputy Dean of Faculty of Computer Systems & Software Engineering
Date : June 13th, 2012.

DEDICATION

I wish to dedicate this to my beloved parents
who have given me love and support all the time.

ACKNOWLEDGEMENT

I am grateful and would like to express my sincere gratitude to my supervisor Dr. Adzhar bin Kamaludin for his germinal ideas, invaluable guidance, continuous encouragement and constant support in making this research possible. He has always impressed me with his outstanding professional conduct, his strong conviction for science, and his belief that a Degree program is only a start of life-long learning experience. I appreciate his consistent support from the first day I applied to post-graduate program to these concluding moments. I am truly grateful for his progressive vision about my training in science, his tolerance for my naïve mistakes, and his commitment to my future career. I am sincerely thankful for the time he spent proofreading and correcting my mistakes.

My sincere thanks to all my classmates and members of the staff of the Faculty of Computer Systems & Software Engineering, UMP, who helped me in many ways and made my stay at UMP pleasant and unforgettable. Many thanks to fellow final year project students for their excellent co-operation, inspirations and supports during this study.

I acknowledge my sincere indebtedness and gratitude to my parents for their love, dream and sacrifice throughout my life. I cannot find the better words to express my appreciation for their devotion, support and faith in my ability to attain my goals. Special thanks should be given to my friends. I would like to acknowledge their comments and suggestions, which were crucial for the successful completion of this study.

ABSTRACT

Ecommerce using Search Engine Optimization (ESEO) is an ecommerce system that focuses on implementing SEO techniques to improve the website's ranking in Search Engine Results Page (SERP). When a new web system is available online, it takes time for the site to be available in SERP. For a site to be noticed by consumers, it is shown that the site's ranking must be on the top area of the SERP, preferably the first page. This is why SEO is important in improving the ranking of a website. One of the approaches is by modifying various elements in the design of the website to effectively increase the relevance of its content to the query of consumers. In ESEO, item information input by user will be used to customize the content generation of the webpages. The optimization works in the form of auto page generation without requiring users' active intervention.

ABSTRAK

Ecommerce dengan menggunakan Optimasi Enjin Pencari (ESEO) ialah sebuah sistem ecommerce yang bertumpu pada mengaplikasikan teknik SEO untuk memperbaik kedudukan laman web dalam Halaman Keputusan Enjin Pencari (SERP). Apabila satu laman web dimuatnaik ke atas internet, ia mengambil masa untuk muncul dalam SERP. Kajian menunjukkan satu laman web mesti berada di kedudukan atas SERP untuk dinampak pengguna, seboleh-bolehnya dalam muka surat pertama. Ini merupakan sebab kenapa SEO penting dalam meningkatkan kedudukan laman web. Salah satu cara adalah dengan mengubahsuai elemen-elemen dalam rekaan laman web untuk meningkatkan kaitan kandungannya dengan carian pengguna. Dalam ESEO, maklumat barangan pengguna akan digunakan untuk menghasilkan laman web. Optimasi ini berfungsi dengan penjanaan laman secara automatik tanpa melibatkan usaha pengguna.

TABLE OF CONTENTS

		Page
STUDENT'S DECLARATION		iii
SUPERVISOR'S DECLARATION		iv
DEDICATION		v
ACKNOWLEDGEMENT		vi
ABSTRACT		vii
ABSTRAK		viii
TABLE OF CONTENTS		ix
LIST OF TABLES		xi
LIST OF FIGURES		xii
CHAPTER 1 INTRODUCTION		
1.1	Background	1
1.2	Problem Statement	2
1.3	Objectives	3
1.4	Scope	3
1.5	Thesis Organization	3
CHAPTER 2 LITERATURE REVIEW		
2.1	Search Engine Optimization (SEO)	4
2.1.1	Search Engine	4
2.1.2	Search Engine Optimization	6
2.2	Methods of Improving Website Ranking	8
2.2.1	Paid Search	8
2.2.2	Search Engine Optimization	11
2.2.3.	Comparisons between Paid Search & SEO	21

2.3 Development Models 27

2.3.1 Iterative & Incremental Development (IID) 27
2.3.2 Rapid Application Development (RAD) 29
2.3.3 Comparisons between IID & RAD 31

CHAPTER 3 METHODOLOGY

3.1 Iterative & Incremental Development 32

3.2 Implementation in Project 33

3.2.1 Planning & Requirement 34
3.2.2 Analysis & Design 35

3.3 Project Requirements 40

3.3.1 Hardware Requirements 40
3.3.2 Software Requirements 41

CHAPTER 4 IMPLEMENTATION

4.1 Search Engine Optimization **Error! Bookmark not defined.**

4.2 Hosting on External Webhost 51

4.3 Google Webmaster Tools 52

CHAPTER 5 RESULTS & CONCLUSSION

5.1 Results 54

5.2 Constraints 58

5.3 Recommendations 58

5.4 Conclusion 59

REFERENCES 60

APPENDIX 63

A Gantt Chart 63

LIST OF TABLES

Table No.	Title	Page
2.1	Comparison between Paid Search & Search Engine Optimization	22
2.2	Methods of Reaching Links on Search Engine Result Page [19]	26
2.3	Comparison between IID & RAD	31
3.1	Hardware Requirements	40
3.2	Software Requirements	41
5.1	List of Recommended Practices by Google and Its Application in ESEO	57

LIST OF FIGURES

Figure No.	Title	Page
2.1	Different Sections of Search Engine Results Page [24]	5
2.2	Visits from Three Major Search Engine Results Page [5]	6
2.3	New AdWords Campaign (Locations and Languages) [25]	9
2.4	New AdWords Campaign (Bidding and Budget) [25]	9
2.5	New AdWords Campaign (Creating an Ad) [25]	10
2.6	Google's Keyword Tool [26]	10
2.7	Google Webmaster Tool (Crawl URL) [29]	12
2.8	Sample of XML Sitemap [13]	13
2.9	Title Tag Appears as Headline on Google SERP [24]	15
2.10	Title Tag of the Wikipedia's Webpage [24]	15
2.11	Description Meta Tag Appears as Snippet on Google SERP [27]	15
2.12	Description Meta Tag of the Pimex's Webpage [27]	15
2.13	Use of Header Tags to Differentiate Weightage of Headline [15]	17
2.14	PageRank of CNET.com [28]	18
2.15	Original PR Formula [2]	19
2.16	Number of Web Results Examined Separated by Type [19]	25
2.17	Iterative and Incremental Development Model [20]	27
2.18	Rapid Application Development Model [23]	29
3.1	Iterative and Incremental Development Model [20]	32
3.2	Flow Chart of Adding Item in ESEO	35

3.3	Dynamic URL (Common Practice)	37
3.4	Static URL (Recommended Practice)	37
3.5	Generic Title Tag (Common Practice)	38
3.6	Unique Title Tag (Recommended Practice)	38
3.7	Generic Image Tag (Common Practice)	39
3.8	Meaningful Image Tag (Recommended Practice)	39
4.1	PHP Coding to Generate Title & Meta Tags	43
4.2	Apache Mod_rewrite Rules	44
4.3	Meaningful Anchor Text & Text-based Navigation Links	45
4.4	XML Sitemap Code	46
4.5	Item List in ESEO	47
4.6	PHP Code to Generate Image Tags	48
4.7	H3 Heading Tag Used for the Page's Heading	49
4.8	Rules in Robots.txt	49
4.9	Facebook Page for ESEO (PeepoYa! Version)	50
4.10	Administration Interface of ESEO Webhost Account	502
4.11	Daily Access Graph (Google Analytics)	50
4.12	Website Traffic Source (Google Analytics)	503

CHAPTER 1

INTRODUCTION

This chapter explains about Ecommerce system using Search Engine Optimization (ESEO) that is developed in this project. The first section describes the background of the project. The second section describes the problem statement of the project. The third section describes the objectives for the project. The fourth section describes the scope for the project. Finally, section five describes the organization of this thesis.

1.1 Background

This project is to create a consumer-to-consumer (C2C) ecommerce system by applying Search Engine Optimization (SEO) process to improve its ranking on Google search engine results page (SERP) search.

A consumer-to-consumer ecommerce system is a platform for consumers to perform trades between each other on the internet. This is different from other categories such as business-to-consumer where it serves as an online retail store for a company or business.

Search Engine Optimization (SEO) is a process consists of both onsite and offsite optimization to improve its accessibility to a search engine. In other words, it helps search engine notice and understands the website. Onsite methods involve both internal design and the content arrangement of a website. Offsite methods

focus on placing backlinks on external websites that attracts users to the intended website.

The implementation of SEO in ESEO requires no extra input from users aside from that needed by an average ecommerce system. The system will generate the user's item display page based on item information input by user and stored in database.

1.2 Problem Statement

The problem of a newly created ecommerce system is that it will rank very low on Search Engine Ranking Page (SERP). This directly impedes its main function of allowing users to perform trades because the website cannot attract sufficient visitors from search engines.

Another problem that an ecommerce system faces is the difficulty to implement SEO techniques due to nature of the system. The main content of an ecommerce system is generated by user input data where dynamic URL (URL that contains programming variables) is needed. However, dynamic URL is not advised in terms of SEO practice because search engines will avoid indexing (include in their search result list) repetitive URL.

Although the use of paid search (a paid advertising service) is another alternative solution, the website owner needs to pay for the service fee depending on the popularity of the keyword being advertised. This project chooses the method of Search Engine Optimization (SEO) which is free and often more preferable by consumers.

1.3 Objectives

The objectives of this system are:

i. To develop a consumer-to-consumer (C2C) ecommerce system
ii. To apply Search Engine Optimization techniques on the system
iii. To test the improvement of SEO on the system using Google Webmaster Tools and conformity of SEO practices using Google Search Engine Optimization Starter Guide

1.4 Scope

The scopes of this project are as follows:

i. The ecommerce system includes item registering and browsing, user management, goods and transaction database.
ii. The SEO methods applied will be onsite improvement and practices recommended by Google will be followed.
iii. The tests will be done on the system before and after the application of SEO to show improvement in related metrics.

1.5 Thesis Organization

This thesis contains five chapters.

Chapter one is the overview of the project. It consists of five parts, including introduction, problem statement, objective, scope and thesis organization.

Chapter two presents review on previously conducted research and studies, including comparison of two methods and development model.

Chapter three explains the model that is used in this project. The process of the development is discussed in detailed here.

Chapter four shows the implementation of this project.

Chapter five then concludes the project and thesis with results.

CHAPTER 2

LITERATURE REVIEW

This chapter reviews and justifies the techniques chosen for this project based on past researches. It consists of four sections: The first section explains briefly about Search Engine Optimization; the second section compares between two techniques associated with Search Engine Optimization; and the third section compares two methodologies which will be adapted into the project.

2.1 Search Engine Optimization (SEO)

2.1.1 Search Engine

The World Wide Web (WWW) is a system of connected documents on the Internet, which contains not only text, but also multimedia files like pictures, audio and video files [1]. The physical locations of the data are stored in datacenters around the world and served to users via webservers. But both the datacenters and webservers are not responsible for the arrangement or indexing of the rapidly growing amount of data, which leaves the task to a search engine that almost directly affects the accessibility of the web contents.

A search engine is a database of Web pages that indexes web pages and provides a way to search the database. Search engines then rely on spiders—software that followed hyperlinks—to find new Web pages to index and insure that pages that have already been indexed are kept up to date [2]. Generally, search

results on search ranking results page (SERP) are then divided into two main categories, the paid search results and organic search results as shown in Figure 2.1 below.

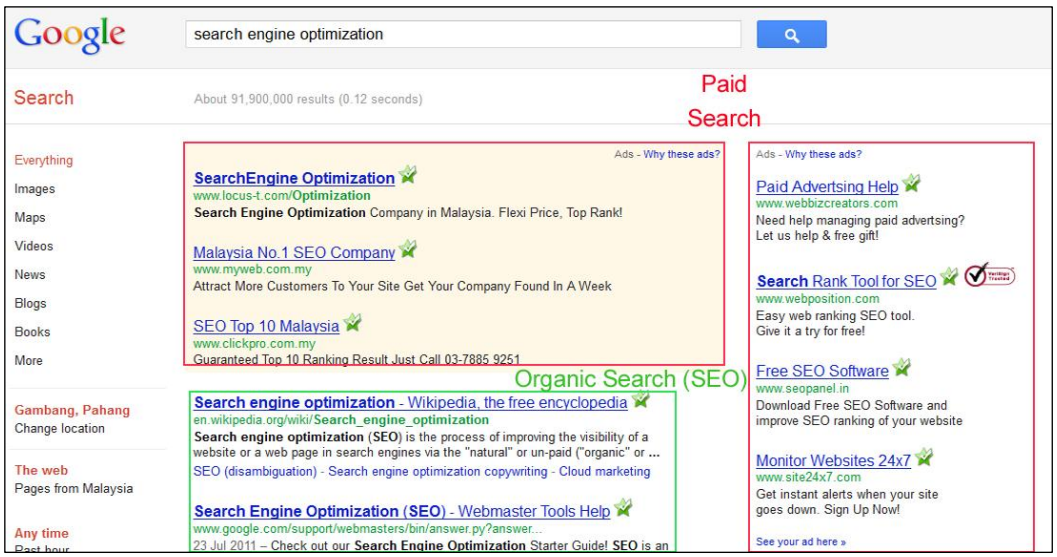


Figure 2.1: Different Sections of Search Engine Results Page [24]

Paid search is more widely known as search engine marketing (SEM). It is an advertising paradigm where marketers purchase small text ads that are triggered when certain keywords are searched [3]. Paid search are often listed on a separate section of a SERP, making it easily distinguishable by users and consumers. This characteristic leads to a further dispute which will then be discussed later in this chapter.

Organic search, often known as search engine optimization (SEO) often involves technical processes where small modifications are made to parts of a website [4]. One noticeable difference between paid search and organic search ranking is that no direct payment is needed for a website to get listed on organic SERP.

2.1.2 Search Engine Optimization

Search Engine Optimization (SEO) is a set of practices to make a webpage easier for search engines to crawl, index and understand its content [4]. As explained, SEO involves making small but important changes to the website, which includes the internal architecture of the website, the content of the website, as well as the effort made to maintain the website. The value a website's search ranking can be depicted below.

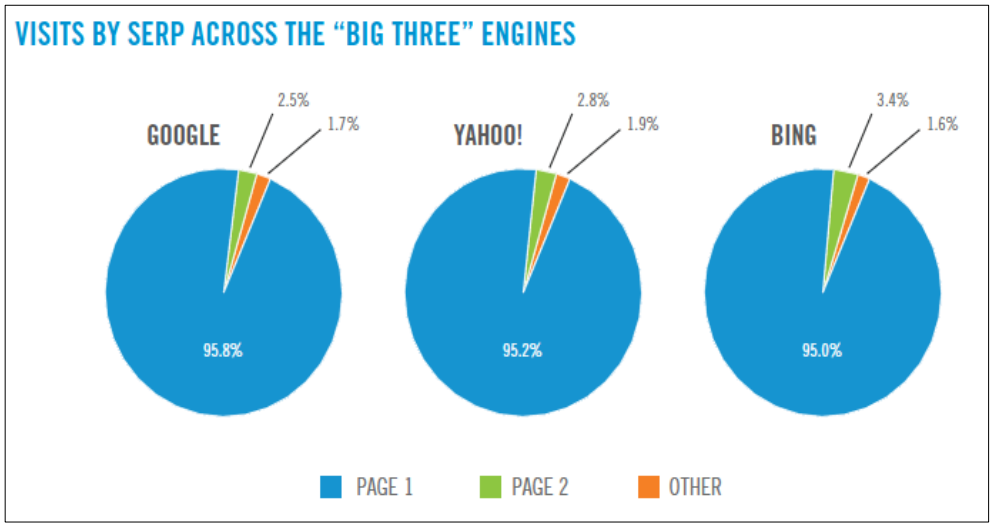


Figure 2.2: Visits from Three Major Search Engine Results Page [5]

According to iCrossing's [5] analysis released on February 2010 as shown in Figure 2.2, websites ranking on second page onwards receive significantly less clicks than websites on the first page. The difference between results of all three main search engines is less than 1%. This shows that without SEO, a website is practically inaccessible by users via search engines.

Conventional SEO techniques or techniques according to guidelines suggested by major search engine involve a set of modifications and practices to follow. While the specific guidelines vary a bit across search engines, they can be

summed up as showing the same content to search engines as you show to users [6].

This is an important characteristic that separates two major approaches of SEO practitioners, commonly known as the black hat practitioners and the white hat practitioners.

According to Google's SEO guide, search engine optimization is about putting a site's best foot forward when it comes to visibility in search engines, but the ultimate consumers are the users, not search engines [4]. Thus, a website with content most relevant to what it's been portrayed to the search engine is more likely to gain ranking using the legitimate way accepted by search engines.

This process includes manipulation of dozens or even hundreds of Web site elements. For example, some of the elements used by the major search engines to determine relevance include, but are not limited to: age of the site, how often new content is added, the ratio of keywords or terms to the total amount of content on the site, and the quality and number of external sites linking to the site [7]. A white hat SEO practitioner aims to promote the website by improving aligning the modifications to the content of the site. In other words, white hat SEO practitioners sell what they show.

In contrast, black hat SEO practitioners employ illegitimate methods to temporarily boost the page rankings. These tactics often involve spamdexing which deliberately manipulates the search engine indexes. An example of is a technique known as scraper site categorized under content spam [8]. This method steals content from its original page and "create" duplicate pages using the content. This can cause either a decline in the ranking of the original page due to duplication of content, a rise in ranking of the duplicated page due to the stolen ranking from the original page, or both.

Black hat SEO techniques may be able to generate quick increase of page ranking in SERP, the increase is usually temporary compared to white hat SEO techniques. On top of that, Search Engine Marketing (SEM) firms that are found to be practicing such techniques will be penalized heavily or banned permanently from search engines. For example, the SEO consulting company Traffic Power was banned from the Google index. In addition, Google also banned Traffic Powers' clients [6].

The risks of getting banned and the ethical issues of employing black hat SEO techniques clearly shows that any web developer should employ the white hat SEO approach. Consequently, the SEO approach employed for this project is the white hat approach.

2.2 Methods of Improving Website Ranking

As discussed earlier, there are currently two main methods of improving a website's accessibility in terms of search engine ranking: (i) Paid search, (ii) Search Engine Optimization.

2.2.1 Paid Search

Generally, paid searches are listed on a separated search ranking list noticeable by users. It requires that advertisers compete for top listing position through bidding in ongoing auctions and then paying when users click on their advertisements, making paid search a flexible and accountable form of advertising. Since its introduction in 1998, paid search has become the dominant form of online advertising and led to Google's $140 billion market capitalization in 2006 [9].

AdWords, Google's main advertising product is one exact example that offers advertisement in terms of paid search.

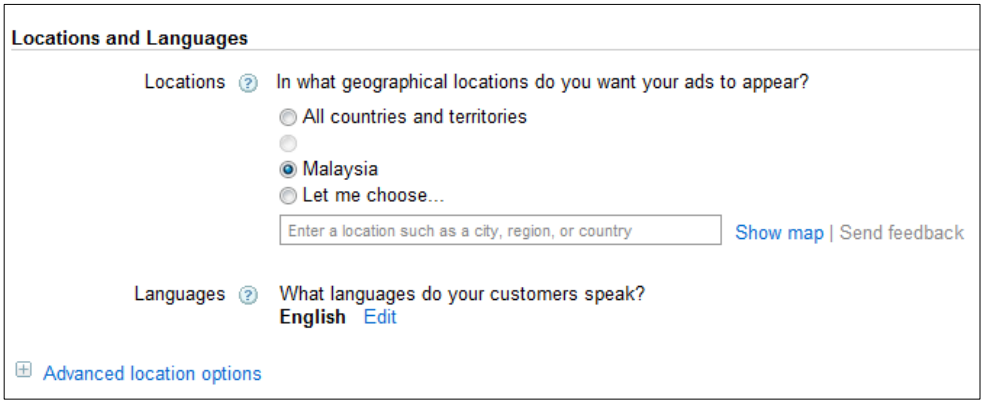


Figure 2.3: New AdWords Campaign (Locations and Languages) [25]

To create a new AdWords campaign, an advertisement owner starts by choosing the location and language of the intended customers as shown in Figure 2.3. This will limit the advertisement to be shown to relevant customers only. For example, the advertisement of a Malaysian restaurant will only be shown to searches originating from Malaysia. Owner may choose to have advertisement shown to global searches too, which in return, increases the scope of customers, as well as the bidding price.

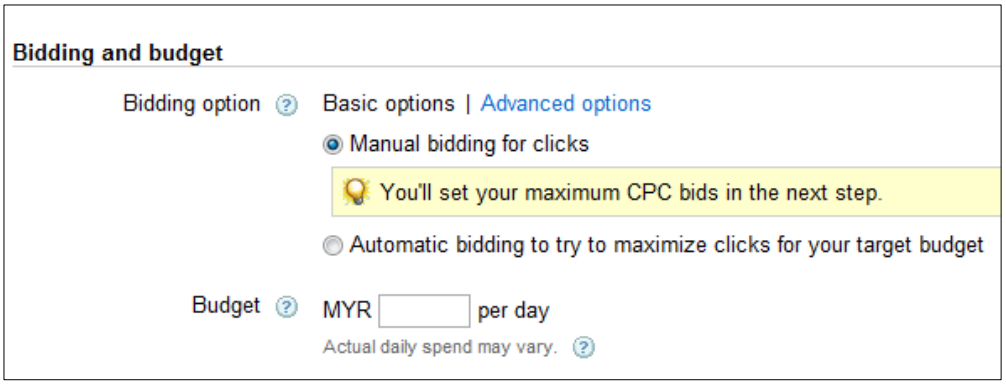


Figure 2.4: New AdWords Campaign (Bidding and Budget) [25]

The bidding and budget is another part that the owner can alter according to his planned budget for the advertisement as shown in Figure 2.4. AdWords will not charge more than the budget set by limiting the number of appearance of the advertisement. For example, by setting RM5 as the budget per day, an

advertisement with Cost-Per-Click (CPC) of RM2 will not appear more than 2 times on the paid search section per day.

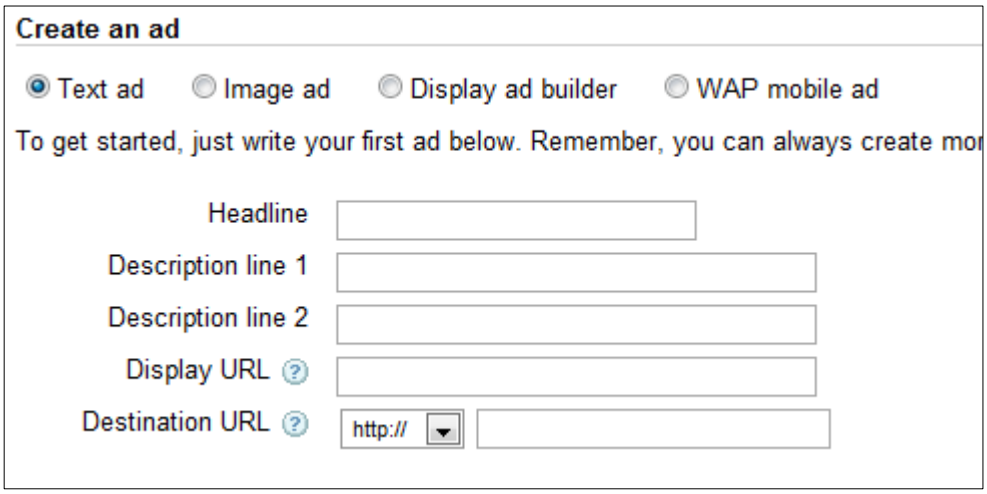
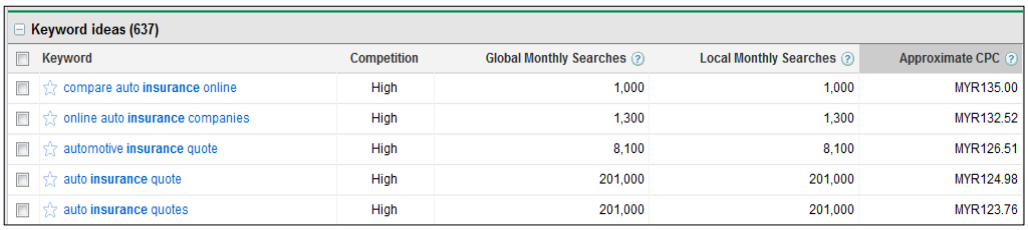


Figure 2.5: New AdWords Campaign (Creating an Ad) [25]

After that, the owner is required to enter several details, including a headline and a short description of his website and the URL of his website which will appear on the paid search ranking section (Figure 2.5).

The advertising keyword(s) is also set in the same page. The keyword must be relevant to the content of owner's website, or it will not rank well even if it's paid search. Each keyword also has different approximate bidding price, depending mainly on the number of competitors currently bidding for the same keyword and the average searches on the keyword. Naturally, popular keywords can have very high Cost-Per-Click (CPC).



Keyword ideas (637)

Keyword	Competition	Global Monthly Searches	Local Monthly Searches	Approximate CPC
compare auto **insurance** online	High	1,000	1,000	MYR135.00
online auto **insurance** companies	High	1,300	1,300	MYR132.52
automotive **insurance** quote	High	8,100	8,100	MYR126.51
auto **insurance** quote	High	201,000	201,000	MYR124.98
auto **insurance** quotes	High	201,000	201,000	MYR123.76

Figure 2.6: Google's Keyword Tool [26]

For example, the keyword ‘compare auto insurance online’ has a CPC of RM135 as shown in Figure 2.6. In other words, a click on the website directed from Google’s search result of the keyword will cost the owner RM135.

Owners are given mainly two models of payment for the advertisement, which are Cost-Per-Click (CPC) and Cost-Per-Impression (CPM). Cost-Per-Click charges the owner each time a valid click is recorded (AdWords has its undisclosed algorithm of detecting spam clicks or unintentional repeated clicks which are defined as invalid clicks [10]). On the other hand, Cost-Per-Impression is the cost per one thousand impressions, where each impression is a single appearance of the advertisement on a web page [11].

After creating an advertisement, an owner should continue to monitor and make changes to his campaign to suit his business preferences. When done correctly, this will certainly make a website appear on top 3 of the paid search listing of specific keywords as long as the owner can afford to pay.

2.2.2 Search Engine Optimization

While paid search ensures good search ranking by charging a bidding price for keywords, Search Engine Optimization (SEO) is a different way of increasing a website’s ranking that requires more technical knowledge of web design and web mastering. In SEO, there are many methods that are used, which can roughly be categorized into three main approaches: (i) indexing, (ii) on-site optimization and (iii) link-building [2].

i. Indexing

Indexing is the process of attracting the search engine spiders to a site, with the goal of getting indexed (included in the search engine’s database) and hopefully

rank well by the search engine quickly [2]. This means that, when a webpage is created, it is not immediately indexed onto any search engine.

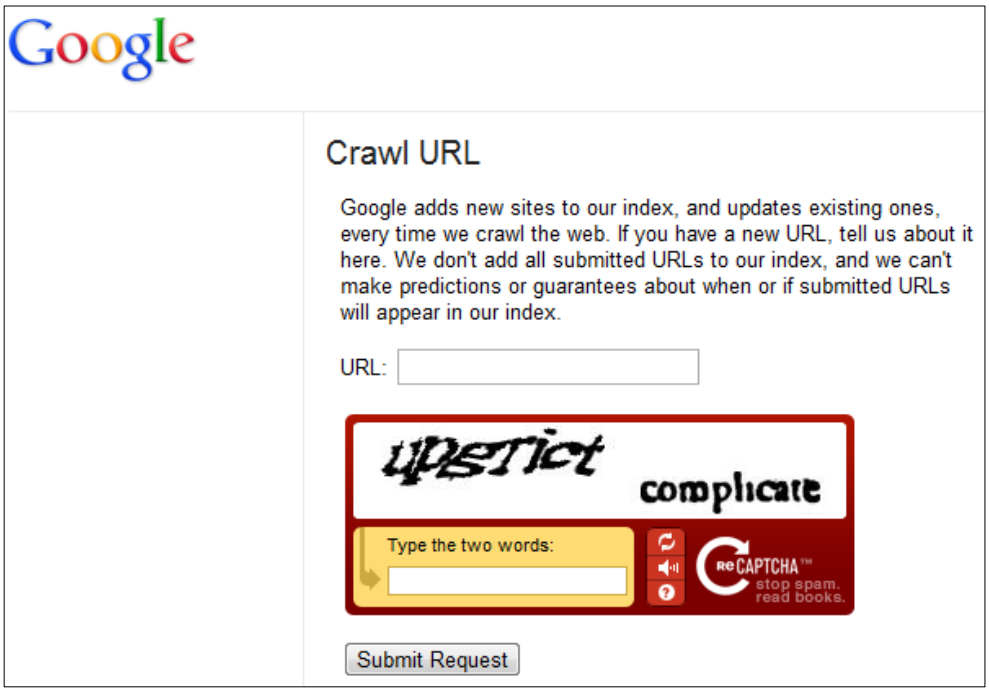


Figure 2.7: Google Webmaster Tool (Crawl URL) [29]

One direct way of getting indexed is by submitting the URL of a new webpage to Google so that crawlers will be sent to crawl and possibly index it (Figure 2.7). Google has warned webmasters not to spam this service by submitting multiple URL directing to the same page as it will not increase the page's ranking and is very likely to get penalized for abusive behavior.

Although this may look like an easy and direct way of getting a new webpage indexed, many SEO practitioner do not encourage a manual site submission through this service as it indirectly means that your site is not well-designed enough to be automatically detected by crawlers [2].

```
<?xml version="1.0" encoding="UTF-8"?>

<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">

   <url>

      <loc>http://www.example.com/</loc>

      <lastmod>2005-01-01</lastmod>

      <changefreq>monthly</changefreq>

      <priority>0.8</priority>

   </url>

</urlset>
```

Figure 2.8: Sample of XML Sitemap [13]

In Figure 2.8, a sample of a website sitemap is shown. The <url> tag contains one URL which the user wishes to include in the sitemap while the various tags inside contains the metadata of the URL, such as <loc> for the URL, <lastmod> for date of last modification, <changefreq> to show how often the URL is updated and <priority> as a scale to show the relative importance of this URL compared to other URL on the same sitemap.

This believe prompts to other methods of getting indexed, one of which is by submitting a sitemap of the website. Another reason is that not all web pages on a website can be reached by crawlers, including links that are dynamically generated using JavaScript, Ajax, and Flash contents and so on, due to the limitation of crawlers [12].

According to Sitemap.org [13], a Sitemap is an XML file that lists URLs for a site along with additional metadata about each URL (when it was last updated, how often it usually changes, and how important it is, relative to other URLs in the site) so that search engines can crawl the site more intelligently. The

URL and metadata in sitemap not only reduce the chance of crawlers missing any hard-to-reach page on a website as mentioned, but also help crawlers understand the structure of a website, further improving its search engine accessibility.

Other than direct submission and creating sitemap, another widely used and effective indexing method is by publishing the link of the web page on user-developed content site. This includes blogs, wikis, social bookmarking sites and the use of RDF Site Summary (RSS) feeds. According to Malaga [2], an optimizer who wants to get a new site indexed quickly can use social bookmarking to bookmark the new site. Since the search engine crawlers visit these sites frequently, they will find and follow the links to the new site and index them. Not to be confused with social networks, social bookmarking sites are more focused on sharing of links and organizing bookmarks. Some popular social bookmarking sites include Delicious, Digg, reddit and so on.

ii. On-Site Optimization

On-site optimization involves more effort on the design of the web pages, including the use of HTML tags, updating and optimizing displayed content as well as formatting.

According to Google [4], <title> tags should contain unique topic for each page on the same website, instead of general words like Homepage or About Us. This is often overlooked by webmasters because they think that title is merely something that appears on the browser tab. In Figure 10, the content of the <title> tag highlighted in red “Search engine optimization – Wikipedia, the free encyclopedia” is used as the headline of that website in Google search as shown in Figure 2.9.

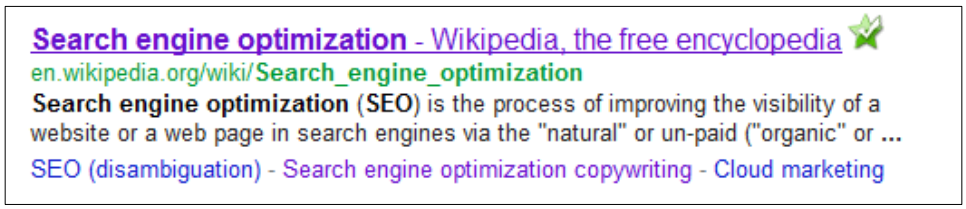


Figure 2.9: Title Tag Appears as Headline on Google SERP [24]

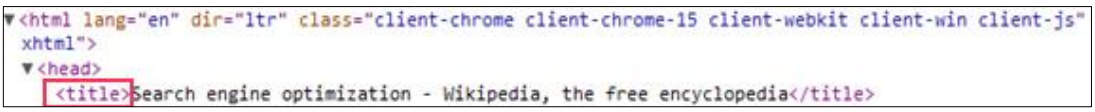


Figure 2.10: Title Tag of the Wikipedia's Webpage [24]

A unique <title> tag helps crawlers to get a good idea of what the page is about and it affects the ranking of the page in Search Engine Result Page (SERP). The <meta> tag helps in a similar way by providing a summarized description of the page content to crawlers. Although it does not affect the page's ranking, it may help crawlers to show relevant snippet on SERP so that users know the content of the website [4]. As shown in Figure 12 below, the content of a <meta> tag is used as the snippet of that website in Google SERP in Figure 2.11.

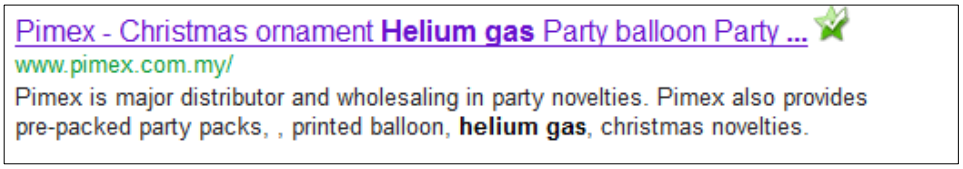


Figure 2.11: Description Meta Tag Appears as Snippet on Google SERP [27]

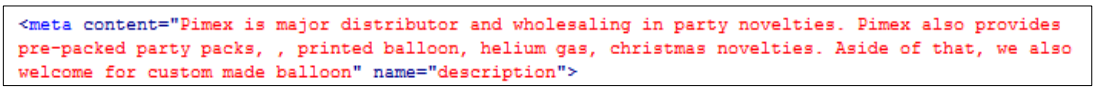


Figure 2.12: Description Meta Tag of the Pimex's Webpage [27]

Other than HTML <title> and <meta> tags, updating and optimizing page content is another way of on-site optimization. This however, involves a greater effort than simple modification of HTML tags.

According to Microsoft's Bing [14], a great way to attract people to your site and to keep them there is to ensure that you have filled your pages with valuable content. This view is further supported by Google in their SEO Guide [4], "Creating compelling and useful content will likely influence your website more than any of the other factors discussed here."

Quality content here mainly emphasizes on three aspects, relevance, originality and freshness. Relevance is how accurate the content presented is, compared to how it appears to search engines. For example, when a website is about body fitness, the content should be related to health and keeping body fitness. If the webmaster chooses to put up some insurance advertisement to earn extra income, visitors will be shocked and very annoyed to find insurance policies after clicking on a search result on body fitness. This will affect the ranking of the website in long run and possibly getting penalized by search engine for deceptive behavior.

Originality of content refers to contents that are unique to the website, instead of duplicated idea or copied text from other places. According to Google [15], Google tries hard to index and show pages with distinct information. In the rare cases in which Google perceives that duplicate content may be shown with intent to manipulate our rankings and deceive our users, we'll also make appropriate adjustments in the indexing and ranking of the sites involved. As a result, the ranking of the site may suffer, or the site might be removed entirely from the Google index, in which case it will no longer appear in search results.

Freshness of content refers to how often a website update its' content where frequently updated websites are more favorable than static pages. This is supported by Steve [3] who wrote, "Google crawls sites that frequently produce quality content more rapidly than they do other, more static properties. The company scours these sites to keep its index fresh by unearthing a trove of new links. Sites that don't publish often are not indexed as rapidly".

After content optimization, formatting is another major part of on-site optimization where various technical aspects of the website design is taken into consideration. Without going deep into the coding of website, the example of a well-structured web page is illustrated in Figure 2.12 below.

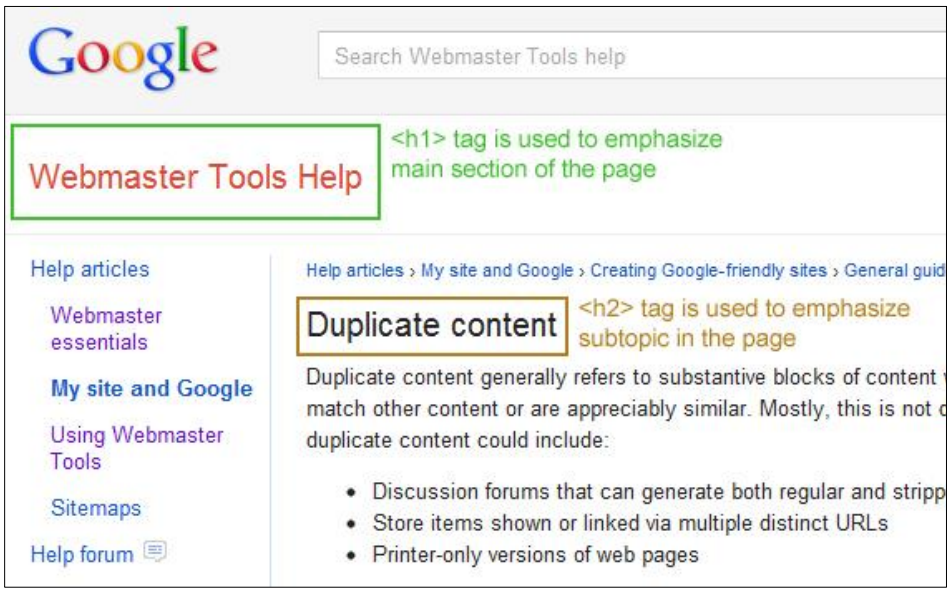


Figure 2.13: Use of Header Tags to Differentiate Weightage of Headline [15]

According to Alan [16], webmaster should imagine that search engine crawlers read like human where contents are read from top to bottom, left to right, including how important headlines are emphasized using different header tags (h1, h2… h6). This also applies for the emphasis tag that emphasize on a stretch of important text.

iii. Link-building

Link-building is the effort of building quality backlinks, which are a major factor in search engine ranking. A backlink is a hyperlink from a site to the target site [2]. Again, there are a few factors that a webmaster should consider before placing a link or requesting for a placement of backlink other than the quantity of backlinks. Figure 2.13 shows the PageRank of the CNET.com.

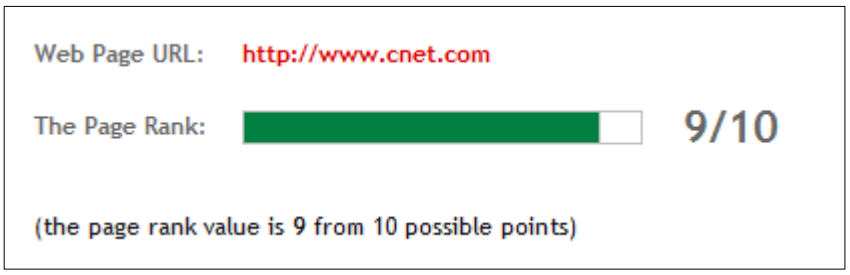


Figure 2.14: PageRank of CNET.com [28]

A basic concept is that Google assigns a Page Rank (PR) to each page in its index where its represented in a number scale from 0 to 10 (10 being the highest) [2]. A backlink placed on a page with high PR has more weight and is more effective than a backlink placed on page with low PR.

For the placement of backlinks, Google advise webmasters to participate in related community and by sharing valuable content which will draw visitors to visit their site and even share the link to their site. Some popular contents include solutions and practical tools that can help visitors.

Webmasters should watch out when they place backlinks to other sites. When the website links to another site with bad reputation, this will affect the website's own reputation. Spamming backlinks will not help either site gain any benefit as well because the PR value will be diluted when more a website places more outbound links.

According to Malaga [2], a basic understanding of PR can be gained by examining the original PR formula (U.S. Patent 6,285,999). Figure 2.15 shows the original PageRank formula.

$$PR(A) = (1 - d) + d(PR(t_1) / C(t_1) + \ldots + PR(t_n) / C(t_n)),$$

Where

'PR' is the Page Rank of a particular Web page,

'$t_1, \ldots, t_n$' are Web pages that link to page 'A',

'C' is the number of links from a page to another page on the Web,

And 'd' is a damping factor.

Figure 2.15: Original PR Formula [2]

In other words, when website A places more outbound links to another website, the PR value will be further divided among the links, causing the outbound links to lose its weight and importance.

Due to the nature of the linking system, webmasters should be careful not to allow links to be placed easily on their site. One way is by using the <nofollow> tag which tells web crawlers not to take the content into account for backlinking purposes. Pages that allow replies or response from visitors commonly use the <nofollow> tag to prevent visitors from spamming backlinks in the field.

On the other hand, a good way to spread backlinks is by sharing content with websites of relevant field [18]. For example, a lawyer who specializes in waterfront issues and riparian rights might share an article with a realtor who specializes in waterfront homes. This will promote a form of mutualism between both parties, where they can link to each other's websites and further strengthen their business with the affiliation.

Other than externally placed links, internal placed links also plays important roles in SEO. Internal placed links are links that direct visitors to pages within the same website. For example, from 'Home' to 'About Us'. A good practice is to use significant anchor text for every hyperlink, instead of pictures and general words like 'Home' or 'Click Here'. This is mentioned in Google's SEO Guide [4], which

explains that crawlers would probably miss links that are not text-based (Using anchor tag), such as Flash-based or JavaScript-based navigation method.

2.2.3. Comparisons between Paid Search & SEO

The comparison will be shown in a table to evaluate their respective strength and weaknesses in different metrics. In this comparison, Google AdWords is used as the example for paid search while SEO covers both SEO firm and in-house SEO.

Metric	Paid Search (Google AdWords)	Search Engine Optimization	Justification
i. Pricing Policies	- Cost-Per-Click or Click-Per-Impression - No monthly fee	- Free (in-house) *or* - According to SEO firms (Outsourced)	- SEO can be cost-efficient by employing staff with SEO skills.
ii. Effective Period	- As long as payment is made - Until outbid by competitors	- As long as SEO effort is maintained	- SEO has long term effect unless webmaster can afford continuous budget for paid search
iii. Time Taken to be Effective	- Immediately after payment	- Usually takes long time to see result	- Paid search will show result much faster than SEO

iv. Skills Required	- Analytical skills to choose keywords and monitor campaign - Marketing skills to ensure profit exceeds investment	- Webmastering skills - SEO skills	- Both requires certain skills for the method to be effective
v. Effectiveness	- Attracts users who already have target in mind	- More preferable by users	- Organic search is likely to attract more visitors compared to sponsored search -
vi. Security	- Constantly being improved by provider, but still vulnerable to 'click fraud' -	- Generally very safe, unless illegitimate SEO approach (black hat) is used	- SEO is less vulnerable to external attack

Table 2.1: Comparison between Paid Search & Search Engine Optimization

i. Pricing Policies

Both methods are considered to have their advantages in terms of pricing, but SEO has a clearer advantage. For paid search, a webmaster does not have to pay for the advertisement to be displayed, but only when clicks have been recorded (for CPC).

As for SEO, the minimum cost can be zero if a company employs webmasters that can perform in-house SEO without contracting the work to external firms. As the cost of keywords bidding can be very high and still rising rapidly due to emergence of competitors for the same keyword and its searching rate, employing permanent SEO solution (staff that can perform in-house or basic automated SEO) can outweigh the profit margin of paid search.

However, this is true and only true if an in-house SEO practitioner is available as contracting an SEO firm can cost varying amount depending on the firm's policy.

ii. Effective Period

In terms of flexibility, paid search campaigns can be terminated anytime according to the webmaster's plan and budget, or until a competitor outbids the webmaster's maximum budget. In other words, as long as the webmaster has the budget, he can stay on top of the paid search ranking list for the desired keywords.

As for SEO, the effect of optimization generally last for very long once a website has topped the SERP. However, continuous maintenance is advised to keep the page content fresh, which can further increase the chance of website staying on top of ranking list. With an in-house SEO practitioner, this is considered an easy effort.

iii. Time Taken to Be Effective

By comparing the time for the results to be shown, paid search clearly outweigh SEO because the effect is almost immediately after payment. On the other hand, SEO is a slow process where the effects can only be seen gradually over a period of time ranging from days to months. If immediate result is a requirement, paid search is the only choice for a webmaster.

iv. Skills Required

For paid search, AdWords is actually designed so that any webmaster without web programming knowledge can use it easily with simple form-filling method. Help notes are also provided along the process to answer doubts that users have. However, users still need to have good analytical skills to monitor their campaigns to ensure that they're optimizing their campaigns for profit. The factors that they have to consider include relevancy of keywords to their field of interest, the price of keywords compared to expected profit and even demographic information of target customers (location and language). All these factors decided whether a campaign is profitable or losing money.

SEO on the other hand requires the webmaster to understand how SEO works in relation to search engine. The main focus is to create a website structure optimized for web crawlers as well as content that can attract clicks and linking.

Both methods require continuous effort as well as domain expertise to a certain degree. However, SEO is considered to be relatively relevant to the domain of webmasters, as compared to the skills required in managing paid search campaigns.

v. Effectiveness

The effectiveness metric of both methods is discussed as the likelihood of a search engine user to click on the search result. In a result of a survey comparing likelihood of user clicking paid search result and organic result (SEO) in six e-commerce Web searching tasks [19], it is shown that organic searches are more likely to be chosen by users (Figure 2.16).

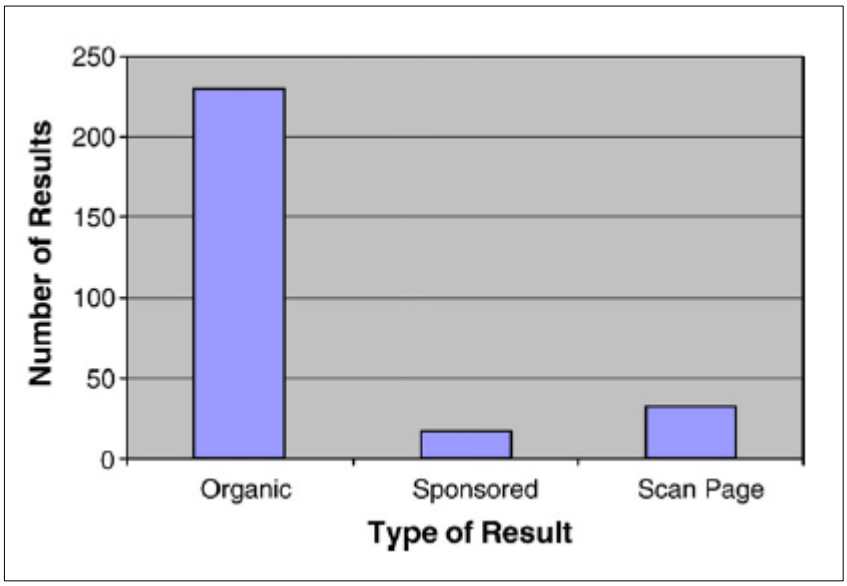


Figure 2.16: Number of Web Results Examined Separated by Type [19]

Methods of viewing links on SERP		
Method	Occurrences	Percentage (%)
Organic by rank	205	73.0
Scanned whole page first	33	11.7
Organic no rank	26	9.3
Sponsored by rank	12	4.3
Sponsored no rank	5	1.8
	281 [a]	100.0

[a] Note: There were 55 tasks where the first action was not logged during a coding error.

Table 2.2: Methods of Reaching Links on Search Engine Result Page [19]

The result shows that participants viewed the organic listings first during 82% of the tasks, compared to 6% for the sponsored listings (Table 2.2). Their hypothesis proved that when using a Web search engine for e-commerce searching, searchers will examine organic links before examining sponsored links. Although there is possibility that this behavior may not apply to every single searcher, the survey provides a good reference to know that paid search may not be as effective as the provider claims.

vi. Security

As mentioned, Google AdWords employ its own algorithm to ensure that its customers are paying for quality service. However, there are still unethical competitors that employ fraudulent methods to exploit the system. In this case, click fraud where competitors click on advertisement just to make the owner pay. This obviously returns no profit for the owner as the clicks are not from possible customers. This method is particularly effective because each click costs the owner a portion of his budget. When his daily budget is reached, his advertisement will not appear on the paid search section anymore, making way for the competitors' advertisements.

This is mentioned by Laffey [9], “It has been estimated that click fraud accounts for 5%–20% of total revenues generated and has grown as paid listing programs have been licensed on a much larger and global basis, making them harder to police.”

On the other hand, websites employing SEO can hardly be targeted by fraudulent activities intended to sabotage them. As most SEO methods can only be applied by the webmaster himself, external intervention is not easy.

2.3 Development Models

2.3.1 Iterative & Incremental Development (IID)

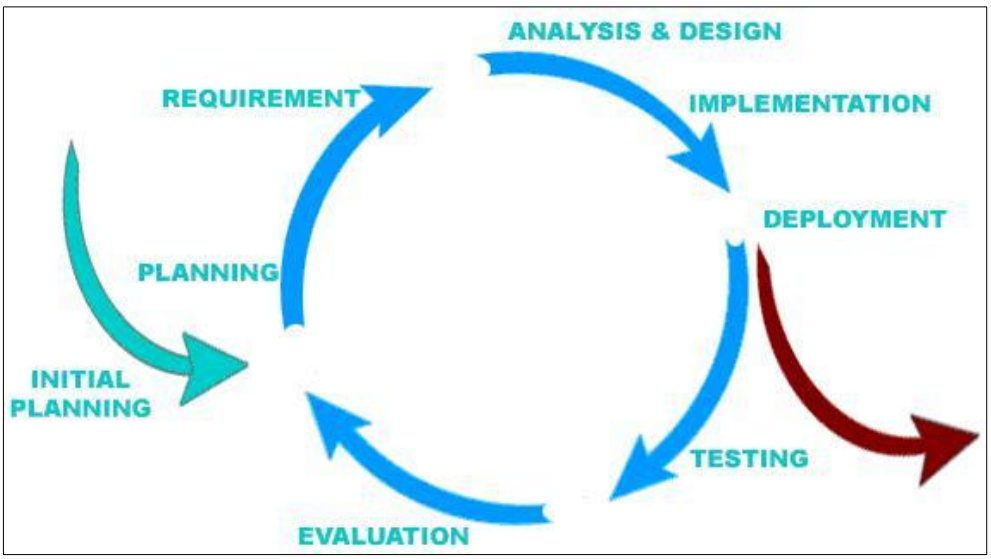


Figure 2.17: Iterative and Incremental Development Model [20]

Iterative and Incremental Development (IID) is a cyclic development process that addresses the weakness of Waterfall model’s rigidness. Starting with the initial planning stage and ending with the deployment stage, IID follows a circular process in between where the stages are repeated as shown in Figure 2.17. The stages in the model

are almost the same as any other development model, except that the middle stages progress in cycles.

Iterative means that the process runs in iterations or cycles where stages are repeated to deal with the complex nature of a software development process. This includes both the requirements stated early in the development and the design made during the development. Iterative allows changes to be made during the development, instead of at the end of the development when everything is almost done. This will save a lot of cost that would otherwise be needed to make changes during final stage of development [21].

Incremental means that the functionalities of the system are added gradually in each iteration, instead of developing everything at the same time. This allows a prototype to be developed earlier and errors or faults can be learned by looking at the prototype [22]. Corrections and changes are then made to the system in the next iteration.

When combined together, IID becomes a more flexible development model that can deal with the robustness of software development as compared to traditional methods.

2.3.2 Rapid Application Development (RAD)

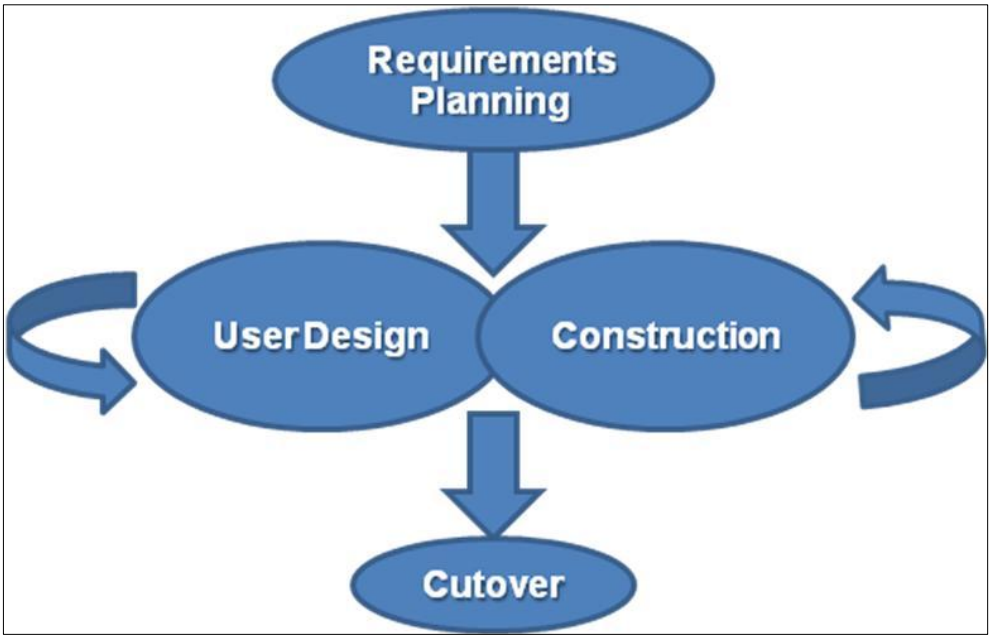


Figure 2.18: Rapid Application Development Model [23]

Rapid application development is another methodology similar to Extreme Programming. It uses minimal planning in favor of rapid prototyping [23].

The main characteristic of this model is that the software development is interleaved with the requirements planning stage as shown in Figure 2.18. A prototype which can include an interface or any part of working software is presented to the customer during the development phase. Customers may then give comments and make changes to the prototype. This process is iterative until customers approve a working model that meets their expectations.

This again, addresses the traditional problems faced in software development model such as the Waterfall model where customers need to wait until the end of the

project to see a working product. This causes large amount of money to be wasted in corrective actions when the product does not meet the requirement of the customer.

RAD's repetitive process in requirements capturing is certainly a key benefit that can save a lot of development cost in terms of requirements changing during later stage of project. Apart from that, customers are also more convinced when they can see something tangible during the project development, instead of waiting for a long period of time without being shown any results.

2.3.3 Comparisons between IID & RAD

Table 2.3 summarizes the comparison between both development models.

Metric	Iterative and Incremental Development	Rapid Application Development
Key Concepts	- Development stages progress in cycles - The system functionalities is added incrementally in each iteration	- Rapid prototyping to present prototypes to customer - Requirements planning is interleaved with prototyping - Reuse of software components
Strengths	- Incremental development allows system to be built in stages - Iterations allow corrections to be made to address each problem	- Close involvement with customer allows better understanding - Rapid prototyping allows user to see a preview of the system and make corrections early
Weaknesses	- May take longer time due to iterations during development	- Quality of system may be affected to the stress on delivery speed - Requires active involvement of customer, fails when customer is unable to cooperate

Table 2.3: Comparison between IID & RAD

CHAPTER 3

METHODOLOGY

This chapter discusses on the methodology chosen for this project as written in Chapter 2. It consists of three sections, including first section that justifies why this methodology is chosen, second section that elaborate on implementation of the methodology in the project and finally, the hardware and software required for the development of this project.

3.1 Iterative & Incremental Development

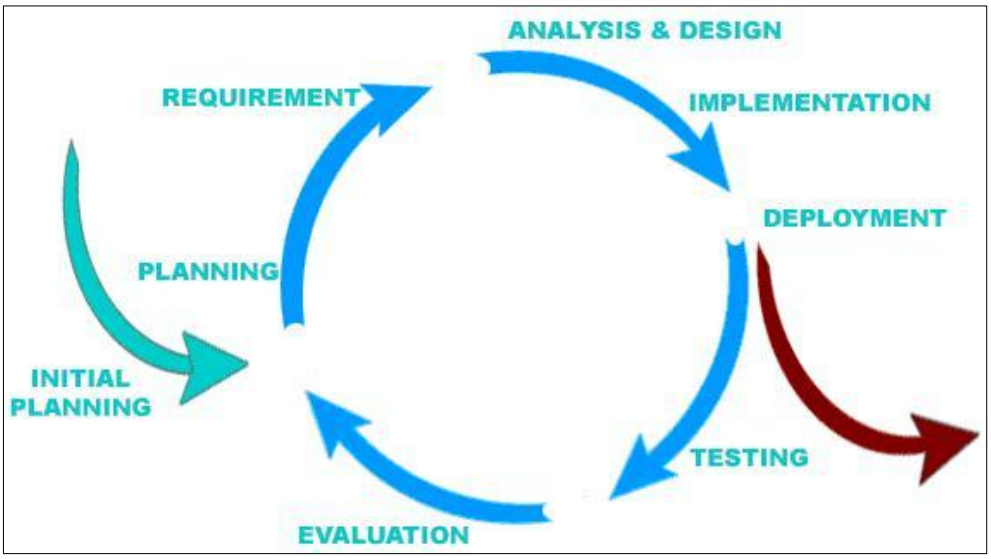


Figure 3.1: Iterative and Incremental Development Model [20]

Figure 3.1 is a diagram showing the stages in Iterative and Incremental Development (IID) model. The strength of this methodology is that development is done incrementally and in in iterations, which is suitable for this project. Incremental means that the system can be developed in parts, where functions and modules are added over time, instead of everything simultaneously [22]. Iteration means that the development process is done in cycles where new changes on requirements or design are made in new iterations. This enables developers to learn from previous iteration and make suitable changes in the current iteration, avoiding changes to be made during implementation. In the implementation of SEO, many features in terms of design need to be taken into consideration. This is why a rigid model such as Waterfall is not suitable for the development of ESEO because the system design may change during development.

3.2 Implementation in Project

The implementation of IID in this project is discussed as followed, separated by stages. Two iterations have been planned for this project where a prototype will be developed and tested in the first iteration and an improved or more complete version will be developed in the second iteration. The reason for planning two iterations is to enable comparison of both versions of systems in terms of the improvement made by the SEO technique.

3.2.1 Planning & Requirement

In the first iteration, the initial requirements for Ecommerce with Search Engine Optimization (ESEO) are determine. As the system is a customer-to-customer e-commerce system, it contains the basic features of an e-commerce site, including user accounts, item listing, shopping cart and purchase summary.

Aside from the normal requirements of an e-commerce system, the additional feature and the focus of this project is the implementation of Search Engine Optimization (SEO) techniques in the system. The SEO features will be the focus of this stage in the second iteration.

3.2.2 Analysis & Design

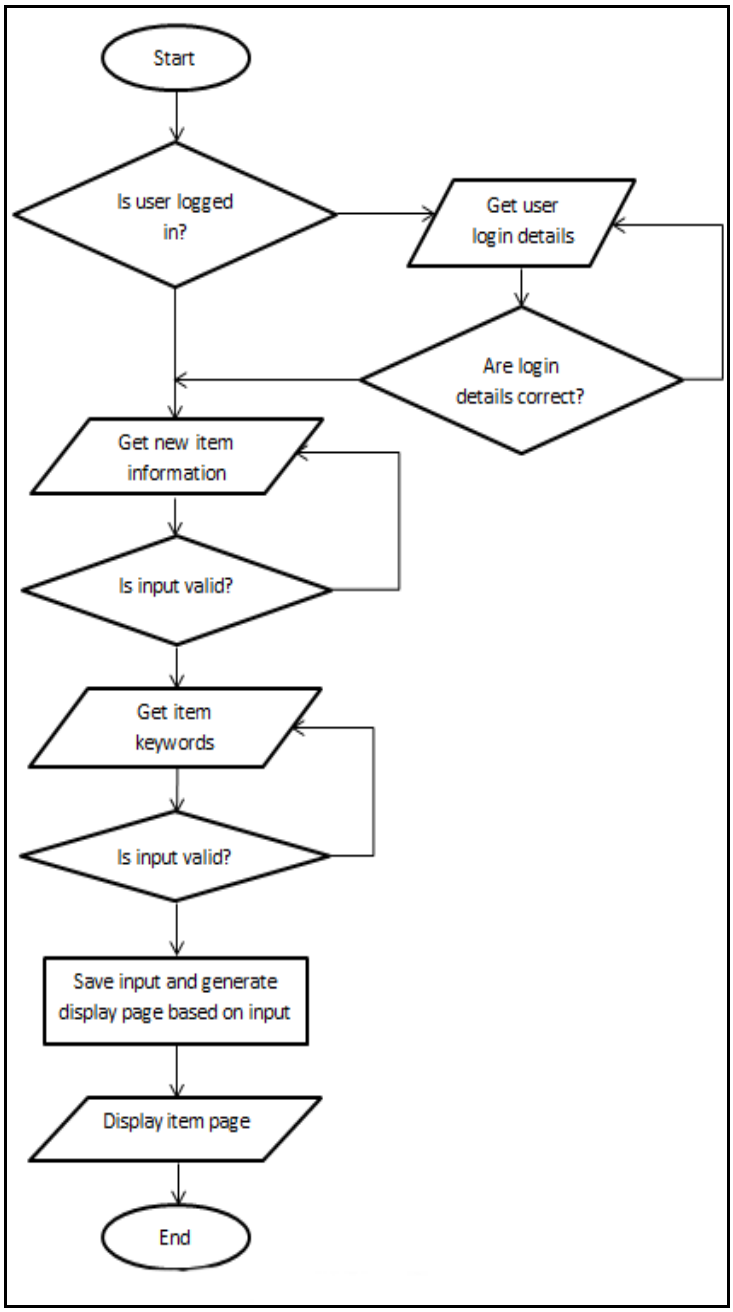


Figure 3.2: Flow Chart of Adding Item in ESEO

The analysis and design stage of this project focuses on two major parts of the system, which are the e-commerce system and the Search Engine Optimization (SEO) feature.

For the SEO feature, the main flow of ESEO is the adding of item by seller. This is because the SEO feature is done by optimizing the item page using the input by user, mainly the keywords that relate to the item being sold.

When a seller chooses to add new item, he needs to be logged in to perform the operation. If not logged in, the user is prompted to enter his login information. Next, the user may enter information of the new item as well as the keyword associated to the item. All required inputs will be validated before the data is saved to the database.

When this is done, the item display page will be generated according to the data by applying SEO approach to optimize the page's accessibility in search engine.

The implementation of SEO in ESEO focuses mainly in auto generation of page coding. By using item information provided by seller, the item page is generated. There are also other practices that do not require user data. Practices that are planned to be implemented in ESEO are:

a. Descriptive and Unique "Title" tags
b. Descriptive and Unique "Meta" Tags
c. URL Containing Words
d. Simple and Hierarchical Directory Structure
e. Text-based Site Navigation Links
f. XML Sitemap for Website
g. Easy-to-Read and Relevant Content
h. Descriptive and Brief Anchor Text

i. Easy-to-Spot Links
j. Descriptive and Brief “Alt” Text for Images
k. Efficient Use of Heading Tags
l. Use of “robots.txt”
m. Use of “nofollow” tag
n. Promote Website the Right Way

Examples of implementation strategies are as followed:

i. URL Rewriting

In SEO practice, it is recommended that a URL contains human-readable words (meaningful to human). This not only provides some information to search engine about the page content, it also helps users to remember the URL easier. Another good practice recommended by SEO is by generating static URL instead of dynamic URL because static URL usually ranks better than dynamic URL.

http://www.somesites.com/forums/thread.php?threadid=12345&sort=date

Figure 3.3: Dynamic URL (Common Practice)

http://www.somesites.com/forums/the-challenges-of-dynamic-urls.htm

Figure 3.4: Static URL (Recommended Practice)

Note how the second URL (Figure 3.4) contains the words “the-challenges-of-dynamic-urls” which can be understood by users instead of the thread id “12345” which means nothing to both a human user and the search engine.

ii. Title, Meta & Header Tags Generation

Title and Header tags often reflect the content of the page. SEO practitioners advise that web designers avoid using generic terms that is likely to reappear on other websites.

```
<title>Homepage</title>
```

Figure 3.5: Generic Title Tag (Common Practice)

```
<title>ESEO Homepage</title>
```

Figure 3.6: Unique Title Tag (Recommended Practice)

In the first example (Figure 3.5), pages with title “Homepage” are all over the Internet and considered repetitive content by search engine. The second example (Figure 3.6) uses the word “ESEO” which is the trademark of the website and is likely to be considered unique.

In ESEO, the item display pages will be generated according to the item information where the mentioned HTML tags will change according to each item stored in database.

iii. Image Tag Generation

One of the limitations of a search engine is that it can only detect text-based content in a webpage. Flash-based content and images practically mean nothing to a search engine and therefore, will not contribute any merit to the page’s ranking. However, the “alt” inside a HTML Image tag can be used by search engine to understand what the image may be.

```
<img alt=”Car” src=http://www.somesite.com/Car  Image1.jpg />
```

Figure 3.7: Generic Image Tag (Common Practice)

```
<img alt=”Red Proton Wira 1995”
src=http://www.somesite.com/Car_Image1.jpg />
```

Figure 3.8: Meaningful Image Tag (Recommended Practice)

The Image tag in Figure 3.8 provides more information than the one in Figure 3.7. ESEO will generate the image tag according to the item.

c. Testing

In the first iteration, the developed ESEO is tested on a running web server. The basic flow of each functionality is tested by the developer as part of white-box testing effort.

In the second iteration, the testing of ESEO will be focused on the improvement on the website's accessibility in terms of search engine result and conformity to Google SEO Practices. This includes using tracking the website's click through rate using Google Webmaster Tool and using checklist in Google Search Engine Optimization Guide.

d. Evaluation

At the end of each iteration, the ranking of ESEO's website on Google's Search Engine Ranking Page (SERP) is tested and compared to see the difference between before and after the implementation of SEO techniques. Google Search Engine

Optimization Guide will be used check the conformity to SEO practices recommended by Google.

e. Deployment

When the system has been tested and evaluated, it will be deployed on an actual web server to ensure that the system is capable of running in a real web environment and with working database.

3.3 Project Requirements

The hardware and software required for the development of ESEO is described as followed.

3.3.1 Hardware Requirements

Table 3.1 shows the software requirements of this project.

No.	Item	Purpose	Specification
1.	Laptop	i. To develop system ii. To run test on system	- Toshiba M300 - Intel Core 2 Duo P7450 (2.13GHz) - 4.00 GB RAM
2.	USB Thumb Drive	i. To store project files for temporary use ii. To keep backup of project files	- Kingston DataTraveler 4GB
3.	Webhost Server	i. To host the system online	- Disk Space 1.5 GB - Bandwidth 100GB

Table 3.1: Hardware Requirements

3.3.2 Software Requirements

Table 3.2 shows the software requirements of this project.

No.	Item	Purpose	Version
1.	Microsoft Windows	i. To serve as the platform of system development	- 7 Ultimate N
2.	Adobe Dreamweaver	i. To develop and compile system coding	- CS5
3.	phpMyAdmin	i. To manage the database of the system	- 3.4.7
4.	Firefox	i. To preview and test the system during development	- 8.0.1
5.	Adobe Photoshop	i. To design graphical related items	- CS4
6.	Microsoft Office	i. For documentation of project	- 2010

Table 3.2: Software Requirements

CHAPTER 4

IMPLEMENTATION

This chapter discusses the implementations and results of the proposed project followed by the constraints and recommendations.

4.1 Search Engine Optimization

The coding implementation of this project focuses mainly on the auto-generation of HTML coding using PHP and MySQL Database. The implementations are separated by each SEO practice applied.

a. *Descriptive and Unique "Title" Tags*
b. *Descriptive and Unique "Meta" Tags*

As one of the recommendations by Google's SEO Guide, it is recommended that a website uses unique <title> and <meta> tags across pages from the same site to assist the search engine in identifying its content. This is especially applied on the item display page where the display page for each item contains different <title> and <meta> tags.

```
$table_name='tbitem';
$table_name2='tbcategory2';
$table_name3='tbcategory1';

$selectItemQuery="SELECT itemName, itemCategory2, FROM $table_name where itemID='$itemID'";
$result=mysql_query($selectItemQuery);
$rowData=mysql_fetch_array($result);
$itemCategory2=$rowData['itemCategory2'];

$selectCategoryQuery="SELECT categoryName2, categoryId1 FROM $table_name2 WHERE categoryId2='$itemCategory2' LIMIT 1";
$result2=mysql_query($selectCategoryQuery);
$rowData2=mysql_fetch_array($result2);
$itemCategoryId1=$rowData2['categoryId2'];

$selectCategoryQuery="SELECT categoryName1 FROM $table_name3 WHERE categoryId1='$itemCategoryId1' LIMIT 1";

$result3=mysql_query($selectCategoryQuery);
$rowData3=mysql_fetch_array($result3);

echo '<title>'.$rowData['itemName'].' | PeepoYa!</title>';  //Part 1
echo '<meta name="description" content="'.$rowData['itemName'].',
'.$rowData2['categoryName2'].', '.$rowData3['categoryName1'].', PeepoYa!">'; // Part 2
echo '<meta name="keywords" content="'.$rowData['itemName'].',
'.$rowData2['categoryName2'].', '.$rowData3['categoryName1'].'">'; //Part 3
```

Figure 4.1: PHP Coding to Generate Title & Meta Tags

Part 1 of the PHP coding generates a <title> tag based on the item name registered, which is different for each item. Part 2 and part 3 generate <meta> tags based on the item name and item categories which describe the type of the item.

c. URL Containing Words

The Apache server module, mod_rewrite is used together with sets of rules that rewrite URL of dynamic pages to show words instead of numbers that are meaningless to human. One of the rules used are shown below:

```
RewriteBase /

RewriteEngine On

RewriteRule item/(.*)-(.*) browseItem.php?id=$1
RewriteRule item/(.*)-(.*)/ browseItem.php?id=$1
```

Figure 4.2: Apache Mod_rewrite Rules

This rule rewrites URL from http://www.peepoya.com/browseItem.php?id=10&name=Nike_Air_Jordan to http://www.peepoya.com/item/10-Nike_Air_Jordan which appears to contain more meaning to users.

d. Simple and Hierarchical Directory Structure

Main pages of ESEO are kept to its uppermost directory to ease traversal of search engine crawlers and to assist search engine in determining the importance of the

page. One example would be: http://www.peepoya.com/browseItem.php where the item browsing page is right after the Home page.

e. *Text-based Site Navigation Links*

h. *Descriptive and Brief Anchor Text*

Anchor texts in ESEO contains minimal number of words, yet able to show the content of the links. This is considered more preferable to search engine crawlers than only using generic words that are found in a lot other websites. The following HTML code in Figure 4.3 is one of this implementation.

```
<div id="menu">
  <ul class="menu">
    <li class="current_page"><a href="#">PeepoYa! Home</a></li>
    <li><a href="browseItem.php">Browse Item</a></li>
    <li><a href="sellItem.php">Sell Item</a></li>
    <li><a href="faq.php">About PeepoYa!</a></li>
  </ul>
</div>
```

Figure 4.3: Meaningful Anchor Text & Text-based Navigation Links

f. XML Sitemap for Website

An XML sitemap of ESEO has been created, containing the pages that are accessible to users. This excludes pages that should not be directly accessible to users, such as the Error 404 page.

```
<url>
  <loc>http://www.peepoya.com/registration</loc>
  <changefreq>weekly</changefreq>
  <priority>0.50</priority>
</url>
<url>
  <loc>http://www.peepoya.com/browse</loc>
  <changefreq>weekly</changefreq>
  <priority>0.80</priority>
</url>
```

Figure 4.4: XML Sitemap Code

Part of the XML sitemap coding is shown n Figure 4.4, where the Homepage and the Browse page are listed. The “changefreq” tag describes how often the page will be updated. This will give an idea for the search engine crawlers how often they should crawl the page to keep an updated index of the page.

The “priority” tag describes the weightage of the page in relation to the Homepage which will guide the search engine crawlers in determining the importance of each page. The index page http://www.peepoya.com always has a priority of 1.00 due to it being the first page accessible by users from search engine. Registration page is given a priority of 0.5 because recurring users are less likely to visit that page. Subsequently, Browse is given higher priority than Registration because it is very likely to be visited by most visitors and should be given higher priority from the crawlers.

g. Easy-to-Read and Relevant Content

Content in ESEO are according to its heading and are separated accordingly. Without using excessive text, pages in ESEO are easy-to-read while having just enough details for user to understand the content. For example, this item list in Figure 4.5 that contains appropriate headings for the details in the list.

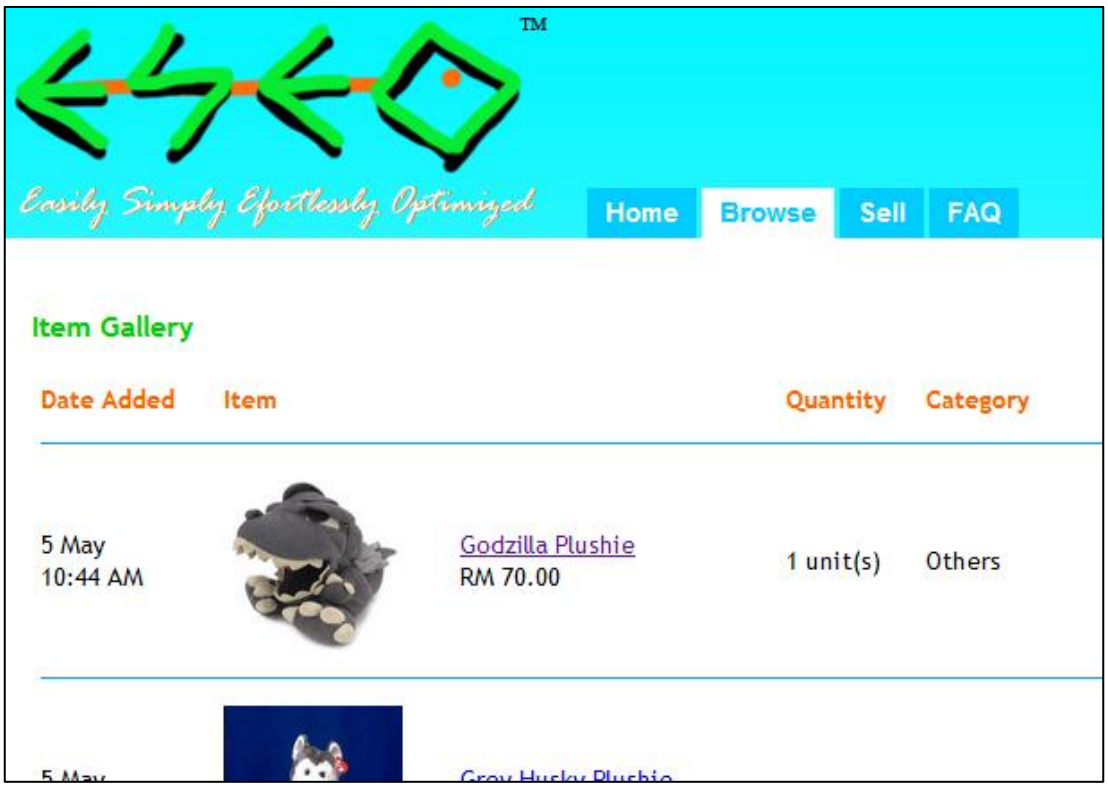


Figure 4.5: Item List in ESEO

i. Easy-to-Spot Links

The links in ESEO are mostly easily noticeable by users without the need of extra effort or attention. Figure 4.5 shows the main menu bar on top which contains 4 sections of the website, each linking to different content. The item names are also underlined, indicating that it is a URL that leads to another page when clicked.

j. Descriptive and Brief "Alt" Text for Images

```
if($rowData['itemImg1']!='-'){
echo '<a href="'.$rowData['itemImg1'].'" target="_blank" rel="nofollow"><img src="'.
$rowData['itemImg1'].'" alt="'.$rowData['itemName'].'" height="180px" width="220px" /></a>';
}
if($rowData['itemImg2']!='-'){
echo '  <a href="'.$rowData['itemImg2'].'" target="_blank" rel="nofollow"><img src="'.
$rowData['itemImg2'].'" alt="'.$rowData['itemName'].'" height="180px" width="220px" /></a>';
}
```

Item name

nofollow

Figure 4.6: PHP Code to Generate Image Tags

In figure 4.6, item names are used for the "alt" attribute of images that correctly describes the images in case images are unavailable. It also allows search engine to identify the content of the image.

m. Use of "nofollow" tag

In Figure 4.6, "nofollow" is used to enclose the URL of the images provided by users so that ESEO's PageRank from will not be affected and the links are not considered as outlink from ESEO.

k. Efficient Use of Heading Tags

In Figure 4.7, heading tags are used for the heading of each page to show significance of the headings. It is not used as an alternative of styling contents.

Figure 4.7: H3 Heading Tag Used for the Page's Heading

l. Use of "robots.txt"

A "robots.txt" file tells search engines whether they can access and crawl parts of a site. It is placed in the root directory of the site. This can prevent the crawler from crawling certain pages because they might not be useful to users if found in a search engine's search results.

```
User-agent: *                  //Line 1
Disallow: /js/                 //Line 2
Disallow: /css/                //Line 3
Disallow: /Error404.php        //Line 4
```

Figure 4.8: Rules in Robots.txt

Figure 4.8 shows some of the rules written in ESEO's "robots.txt". Line 1 defines the crawler that "robots.txt" should be applied on, "*" means that this applies to all available crawlers. Line 2 and line 3 disallow crawlers from crawling the directory

/js/ and /css/ which contain javascript and css files which are irrelevant to users. Line 4 disallows the page Error404 from being crawled because the page should not be accessible to users under normal situations.

n. Promote Website the Right Way

A Facebook page has been created for ESEO as an effort to promote ESEO in the social media site. Social media sites are considered a good form of promotion because of its large amount of audience and it does not impose any fees.

Figure 4.9: Facebook Page for ESEO (PeepoYa! Version)

It is also a good channel for users to promote the website when they like or enjoy the website. If done correctly, this is a very effective form of free advertising.

4.2 Hosting on External Webhost Server

The principles of SEO revolve around building up reputation and ranking of the website. Thus, having a unique domain name that reflects the brand of the website is crucial for this project.

For the purpose of this project, two domain names and hosting accounts have been purchased from Exabytes (www.exabytes.com.my) for two versions of ESEO:

i. www.psmeseo.com

The normal version of ESEO without SEO implementations and with common mistakes made by site owners. The domain name "psmeseo" uses initials that do not mean anything to users.

ii. www.peepoya.com

The version of ESEO improved with SEO techniques. The domain name reflects a unique brand name for the ecommerce business which is a determining factor in advertising the business. Domain name with words is also easier to remember than name that contains long initials.

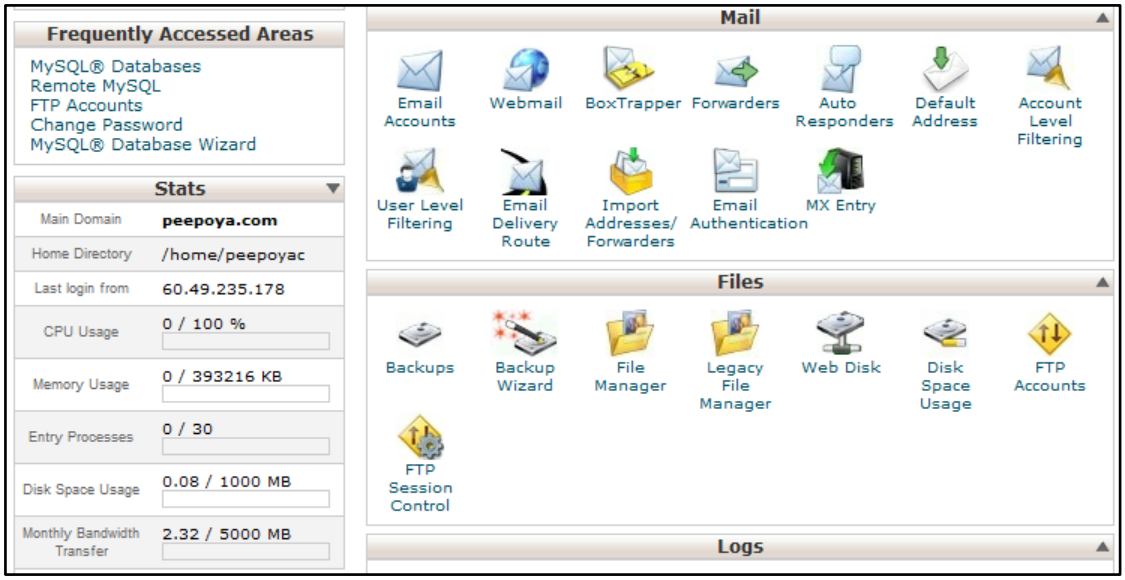


Figure 4.10: Administration Interface of ESEO Webhost Account

Figure 4.10 shows part of the administration page of the webhosting site for the domain www.peepoya.com. Disk and bandwidth usage as well as many other metrics can be observed here. The access logs and graphs also provide useful information for webmasters to analyze the traffic of their website so that they can make appropriate improvements.

4.3 Google Webmaster Tools

Google webmaster tools were also explored in this project. Both Google Webmaster Tool and Google Analytics provide useful analysis functions that can analyze the performance of a website, including the traffic and usage.

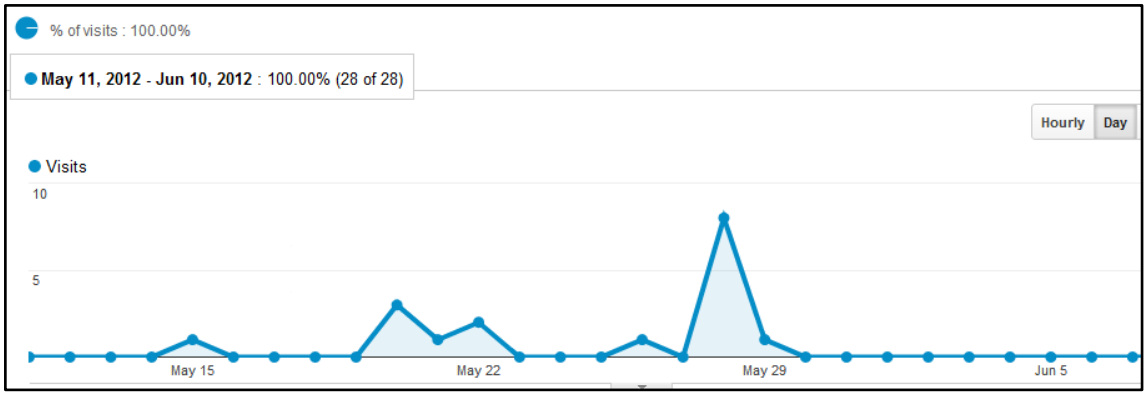


Figure 4.11: Daily Access Graph (Google Analytics)

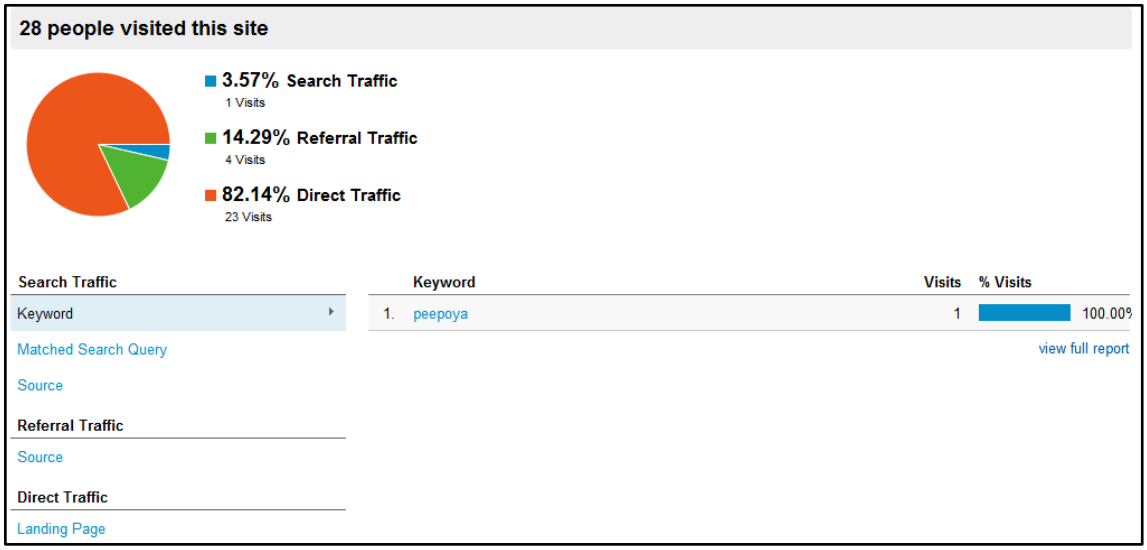


Figure 4.12: Website Traffic Source (Google Analytics)

Figure 4.11 shows the daily traffic graph for "www.peepoya.com" while Figure 4.12 analyzes and separates the different source of traffic. Both functions are extremely useful for webmaster to determine the source of the website's traffic and make suitable effort to increase the website's traffic.

CHAPTER 5

RESULTS & CONCLUSSION

This chapter discusses the results of the proposed project followed by the constraints, recommendations and conclusion.

5.1 Results

The results according to the objectives of this system are as followed:

i. To develop a consumer-to-consumer (C2C) ecommerce system

Two versions of ESEO have been developed, which are hosted on working external file hosting sites. The first version is hosted on the domain name, www.psmeseo.com while the second version is hosted on the domain, www.peepoya.com. Although both versions contain almost identical functionalities, Peepoya is the improved version that contains elements of SEO techniques and practices.

ii. To apply Search Engine Optimization techniques on the system

Search Engine Optimization (SEO) practices have been applied on the second version hosted on www.peepoya.com and the practices conform to the SEO practices recommended by Google which will be explained below.

iii. To test the improvement of SEO on the system using Google Webmaster Tools and conformity of SEO practices using Google Search Engine Optimization Starter Guide

ESEO conforms to the recommended practices by Google Search Engine Optimization Guide. The practices and the results of its implementation are summarized as below:

Recommended Practice	Result (Application)
a. Descriptive and Unique "Title" tags	PHP code is used to retrieve item information from database for generation of unique tags <title> and <meta> tags that describe the item pages.
b. Descriptive and Unique "Meta" Tags	
c. URL Containing Words	Apache module, mod_rewrite is used together with sets of rules that rewrite URL of dynamic pages to show words instead of numbers that are meaningless to human.
d. Simple and Hierarchical Directory Structure	Main pages are kept to the highest possible directory, right after "root" level to show the importance of the pages.
e. Text-based Site Navigation Links	No flash or javascript generated or picture-based links have been used to create navigation links so that search engine can crawl the links without problem.
f. XML Sitemap for Website	An XML sitemap has been created and uploaded onto the server.
g. Easy-to-Read and Relevant Content	Pages contain only minimum amount of text which are relevant to the purpose of the page. Important information of items listed are also arranged neatly to ease users in reading the page.

h. Descriptive and Brief Anchor Text	Meaningful and concise anchor texts are used in links where users understand where the link leads to.
i. Easy-to-Spot Links	Main navigation links are on a menu bar near the top of the pages; Individual item links are underlined in the item list page. This allows users to differentiate links from normal text easily.
j. Descriptive and Brief "Alt" Text for Images	Item names are used for the "alt" attribute of images that correctly describes the images in case images are unavailable. It also allows search engine to identify the content of the image.
k. Efficient Use of Heading Tags	Heading tags are used for the heading of each page to show significance of the headings.
l. Use of "robots.txt"	Robots.txt is used to restrict search engine from crawling and listing irrelevant items in search engine results.
m. Use of "nofollow" tag	"nofollow" is included in image links and email links where users can input their own URL to avoid ESEO's PageRank from being affected.
n. Promote Website the Right Way	A facebook page has been created to promote this website. According to Google's guidelines, social media sites are a good way to promote and attract visits from users.

Table 5.1: List of Recommended Practices by Google and Its Application in ESEO

5.2 Constraints

Although the SERP ranking of the page is expected to increase by applying SEO, the result may differ due to various factors:

i. Search Engine considers a large number of factors when calculating the page ranking in which all of them are undisclosed and therefore, may be left out from this website.

ii. Insufficient user access to the website causing lack of access rate, which is bad for the website's popularity rating.

iii. External factors such as competition from other websites that has similar keywords may in turn lower the ranking of this website, thus nullifying the expected result.

iv. Inability to keep the page updated is a factor that may lower a website's ranking due to inactivity and this is not covered in the scope of the project.

5.3 Recommendations

For future improvement, the website can be constantly maintained and updated by a webmaster to keep its content fresh. This will appeal to search engine as an active website containing fresh content.

User access rate is also another important factor that can be improved on, in order to increase the website ranking. Unfortunately, this factor cannot be covered by means of coding and internal system design. It can be done by various promotion or advertising efforts, such as via social networks which has large number of visitors.

Last but not least, the web design should conform to the latest practices of SEO in order to keep up with the updates of search engines. As search engine providers constantly improve their algorithm to increase accuracy and relevancy of search

results, the factors are constantly being adjusted (added or removed) to produce the best combination. As such, webmasters need to be active in the SEO community to keep up with the changes and apply them to the website.

5.4 Conclusion

In conclusion, this project has achieved most of its objectives. By applying the SEO practices recommended by Google's Search Engine Optimization, a webmaster is practicing correct ways of improving the website's accessibility to Google's Search Engine. Although Google do not directly disclose its complete algorithm in determining website rankings on its Search Engine Result Page, the practices recommended by Google is expected to help webmasters.

Compared to paid advertising, SEO is a long term effort that will eventually pay off when the correct practices are followed. Although SEO usually takes longer time to show improvement in terms of website ranking especially for popular keywords, it is still a method worth considering for website owners who wish to build up the website's reputation and long term popularity. Another benefit is that when practiced in-house by the website owner, SEO can save a lot of money compared to paid advertising.

Due to limited time frame of this project, the potential of SEO may not have been fully demonstrated in this project. If given enough time for the popularity to grow, SEO is expected to show even better results.

REFERENCES

[1] *the World Wide Web*, http://dictionary.cambridge.org/dictionary/british/the-world-wide-web?q=world+wide+web Retrieved on 19th October, 2011.

[2] Ross A. Malaga. *Advances in Computers, Volume 78. Improving the Web.* Elsevier, 2010.

[3] Steve Rubel. *Search Engine Visibility*. Edelman Digital, 2009.

[4] *Search Engine Optimization Starter Guide*, http://www.google.co.jp/intl/en/webmasters/docs/search-engine-optimization-starter-guide.pdf Retrieved 7th October, 2011.

[5] Colin Cornwell. *The Importance of Page-one Visibility*. iCrossing, 2010.

[6] Ross A. Malaga. *Worst Practices in Search Engine Optimization*. In Communications of the ACM - Surviving the data deluge, Volume 51 Issue 12, pp. 147-150. ACM New York, USA, December 2008.

[7] H.H. Clark, S.E. Brennan. *Grounding in Communication*. In Perspectives on Socially Shared Cognition, pp. 127–149. American Psychological Association, Washington, DC, 1991.

[8] *Spamdexing*, http://en.wikipedia.org/wiki/Spamdexing Retrieved on 20th October, 2011.

[9] Des Laffey. *Paid search: The innovation that changed the Web*. Business Horizons, Volume 50, Issue 3, pp. 211-218. Elsevier, March 2007.

[10] *Does increased traffic mean I'm getting invalid clicks?* http://adwords.google.com/support/aw/bin/answer.py?hl=en&answer=6424 Retrieved on 25th October, 2011.

[11] *Cost per impression*, http://en.wikipedia.org/wiki/Cost_per_impression Retrieved on 25th October, 2011.

[12] *About Sitemaps*, http://www.google.com/support/webmasters/bin/answer.py?answer=156184 Retrieved on 26th October, 2011.

[13] *What are Sitemaps?* http://www.sitemaps.org/index.php Retrieved on 26th October, 2011.

[14] *Guidelines for successful indexing*, http://onlinehelp.microsoft.com/en-us/bing/hh204434.aspx Retrieved on 26th October, 2011.

[15] *Duplicate content*, http://www.google.com/support/webmasters/bin/answer.py?answer=66359 Retrieved on 27th October, 2011.

[16] Alan K'necht. *SEO and Your Web Site*. Digital Web, 2004. http://www.digital-web.com/articles/seo_and_your_web_site/ Retrieved on 27th October, 2011.

[17] Kaspar Szymanski. *Quality links to your site*. Google, 2010. http://googlewebmastercentral.blogspot.com/2010/06/quality-links-to-your-site.html Retrieved on 28th October, 2011.

[18] Jim Lodicho. *7 Ways to Improve Your Blog SEO with Inbound Links.* Social Media Examiner, 2010. http://www.socialmediaexaminer.com/7-ways-to-improve-your-blog-seo-with-inbound-links/ Retrieved on 28th October, 2011.

[19] Bernard J. Jansen, Anna Brown, Marc Resnick. *Factors relating to the decision to click on a sponsored link.* In Decision Support Systems, Volume 44 Issue 1, pp. 46-59. Elsevier, November 2007.

[20] *Agile Methodology - Changing Ways of Software Development*, http://www.otssolutions.com/blog/?p=62 Retrieved on 23th September, 2011.

[21] *Using both incremental and iterative development*, http://alistair.cockburn.us/Using+both+incremental+and+iterative+development Retrieved on 24th September, 2011.

[22] *Iterative and incremental development*, http://en.wikipedia.org/wiki/Iterative_and_incremental_development Retrieved on 29th October, 2011.

[23] *Rapid application development*, http://en.wikipedia.org/wiki/Rapid_application_development Retrieved on 29th October, 2011.

[24] *Google search result* (search engine optimization), http://www.google.com.my/#hl=en&source=hp&q=search+engine+optimization&oq=search+en&aq=1&aqi=g10&aql=&gs_sm=e&gs_upl=1776l3895l0l5561l9l7l0l0l0l0l457l1499l2.2.1.1.1l7l0&bav=on.2,or.r_gc.r_pw.,cf.osb&fp=ed46adf8d8a7f467&biw=1280&bih=697
Retrieved on 13th November, 2011.

[25] *Google AdWords*, https://adwords.google.com Retrieved on 25th October, 2011.

[26] *Google AdWords Keyword Tool*, https://adwords.google.com/o/Targeting/Explorer?__c=1201654524&__u=1854972804&__o=kt&ideaRequestType=KEYWORD_IDEAS#search.none
Retrieved on 25th October, 2011.

[27] *Google search result* (helium gas), http://www.google.com.my/#hl=en&q=helium+gas&oq=helium+gas&aq=f&aqi=g10&aql=&gs_sm=e&gs_upl=6005854l6007183l0l6007311l10l9l0l0l0l0l295l1753l0.3.5l8l0&bav=on.2,or.r_gc.r_pw.,cf.osb&fp=ed46adf8d8a7f467&biw=1280&bih=697
Retrieved on 25th October, 2011.

[28] *PageRank Checker* (http://cnet.com), http://www.prchecker.info/check_page_rank.php Retrieved on 28th October, 2011.

[29] *Google Crawler*, https://www.google.com/webmasters/tools/submit-url 28th October, 2011.

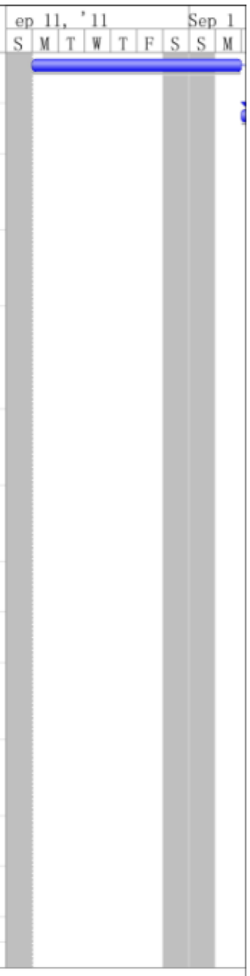
[30] *Google Webmaster Tool*, https://www.google.com/webmasters/tools/dashboard?hl=en&siteUrl=http%3A%2F%2Fwww.mastroharmonica.com%2F Accessed on 14th December, 2011.

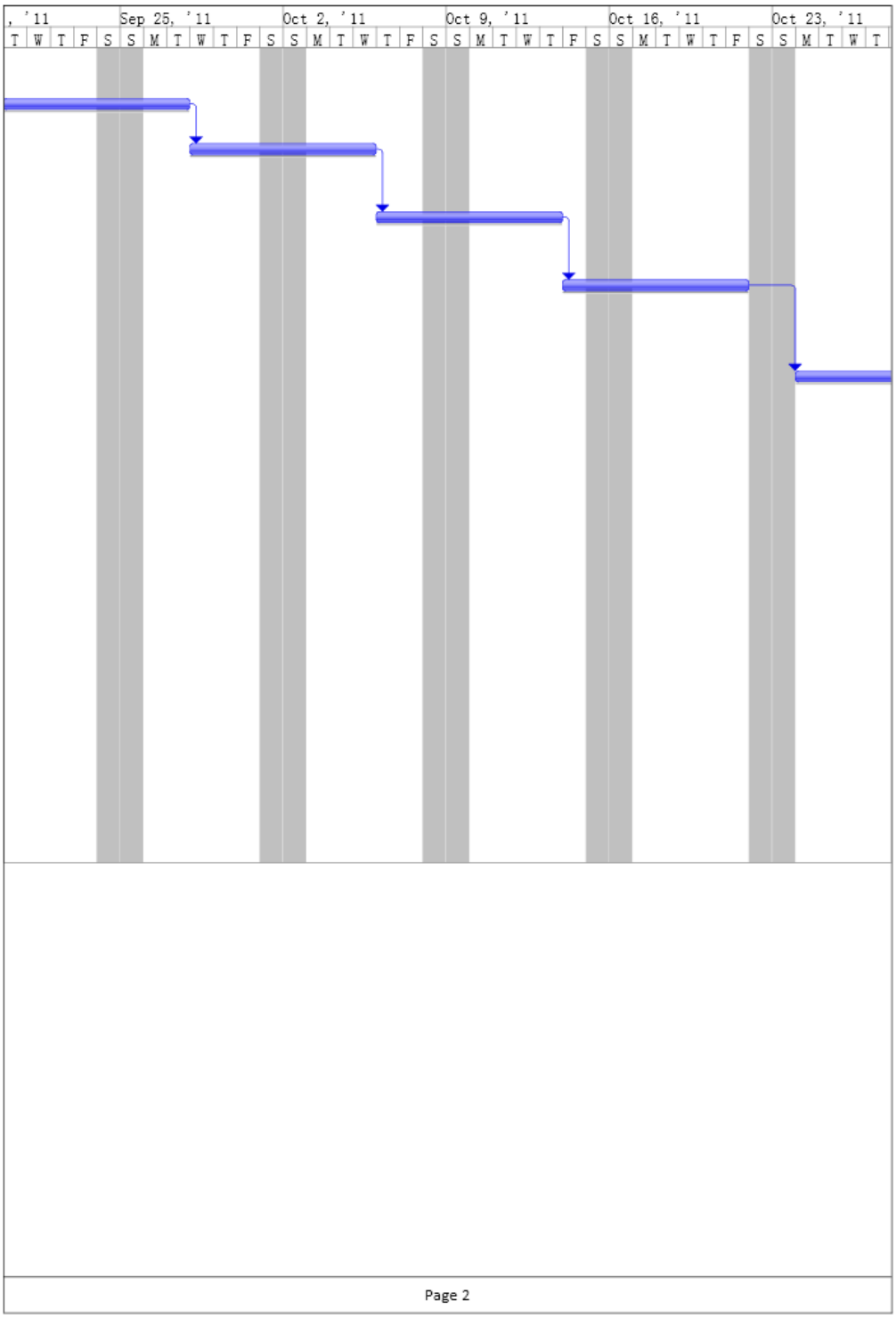
APPENDIX A

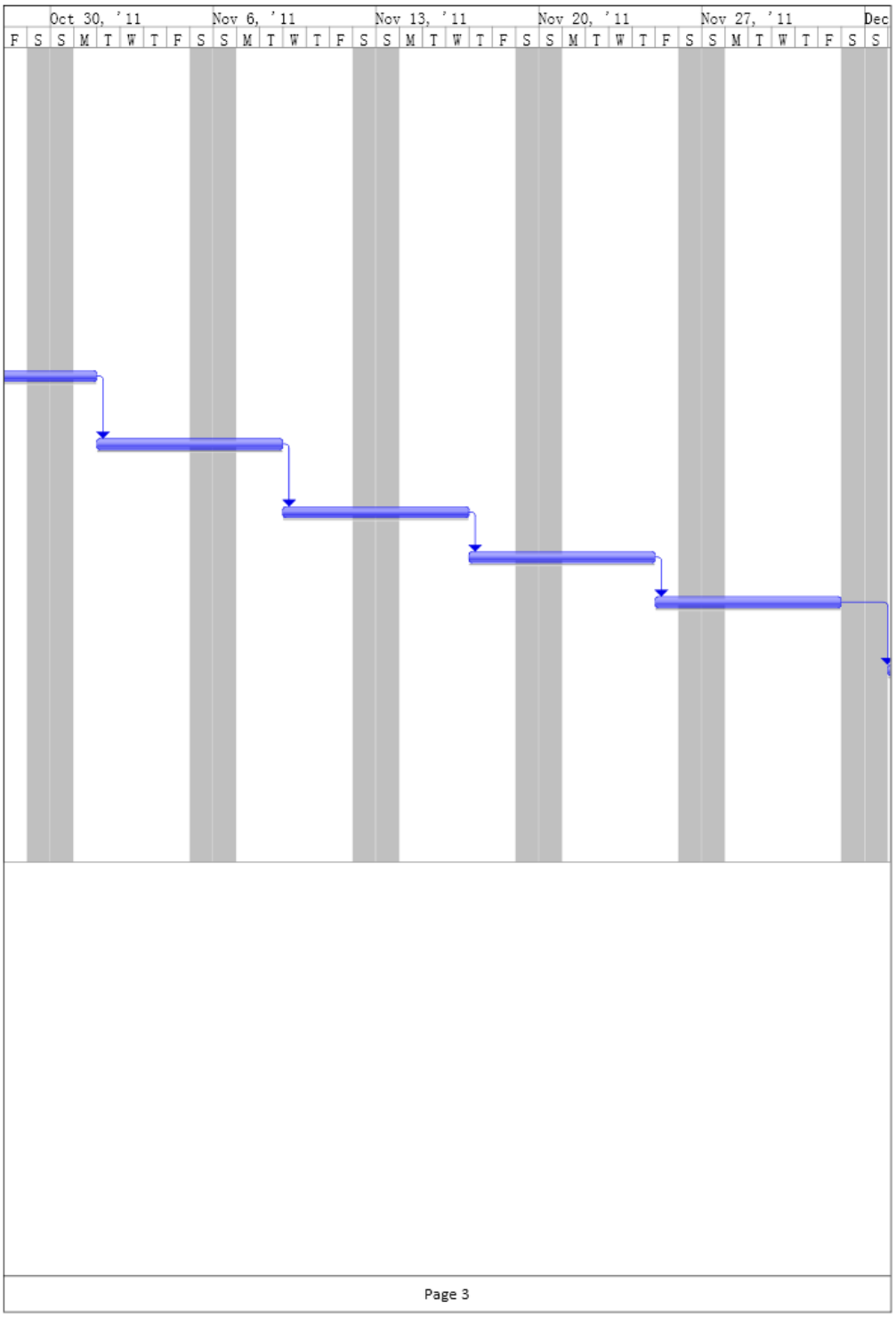
Gantt Chart

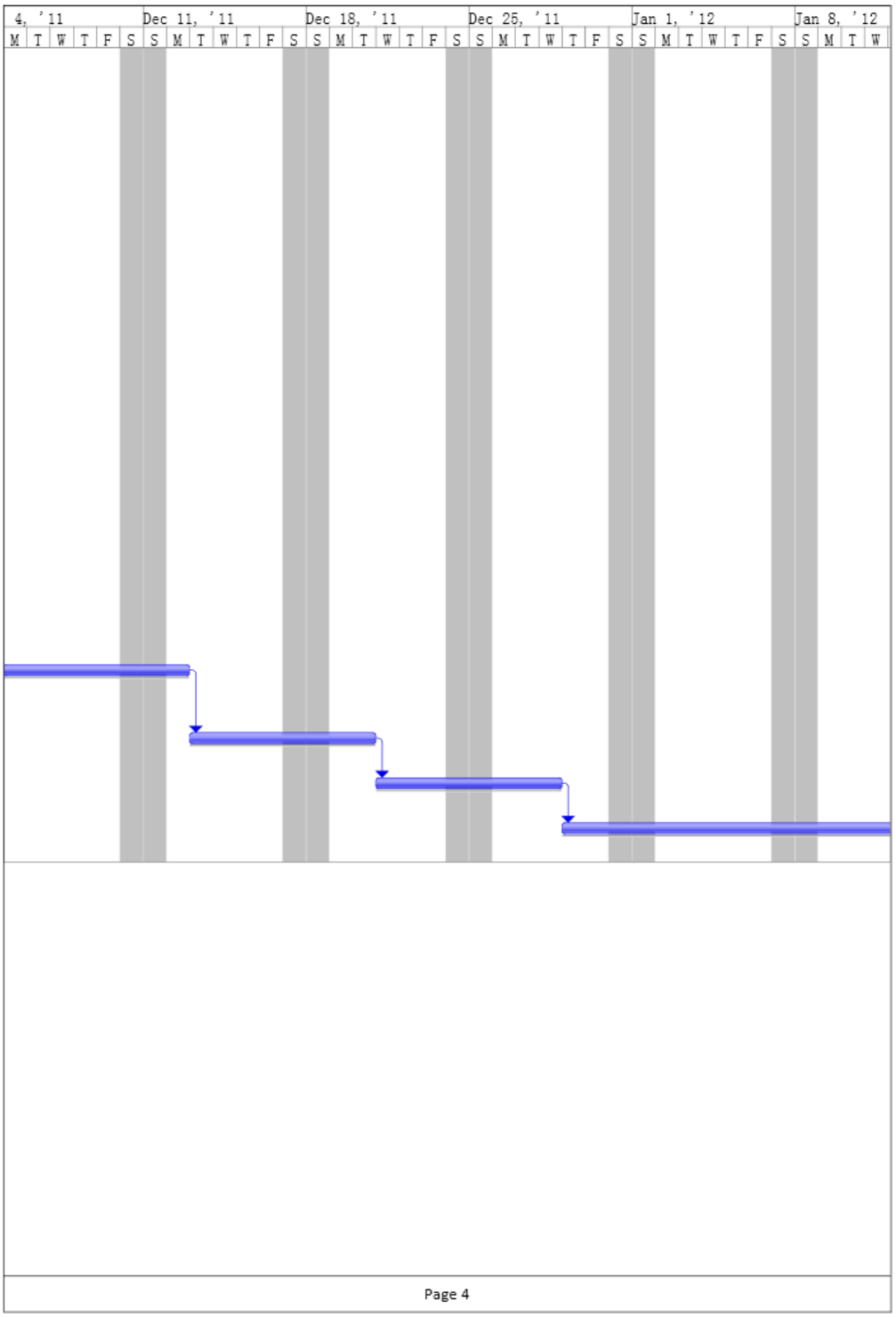
	Task Name	Duration	Start	Finish
1	First Meeting with Supervisor	6 days	Mon 9/12/11	Mon 9/19/11
2	Prepare and Summit Chapter 1	6 days	Tue 9/20/11	Tue 9/27/11
3	Discuss About Title and Possible Alternative	6 days	Wed 9/28/11	Wed 10/5/11
4	Confirm Change of Title, Discuss About Correction	6 days	Thu 10/6/11	Thu 10/13/11
5	Discuss About Literature Review, Submit Literature Review I	6 days	Fri 10/14/1	Fri 10/21/11
6	Continue with Literature Review Part II	6 days	Mon 10/24/1	Mon 10/31/11
7	Submit Part II of Literature Review	6 days	Tue 11/1/11	Tue 11/8/11
8	Correction on Literature Review	6 days	Wed 11/9/11	Wed 11/16/11
9	Complete Literature Review	6 days	Thu 11/17/1	Thu 11/24/11
10	Proceed with Chapter 3 Methodology	6 days	Fri 11/25/1	Fri 12/2/11
11	Submit Correction of Chapter 2, Submit Chapter 3	6 days	Mon 12/5/11	Mon 12/12/11
12	Discuss About Project Overview	6 days	Tue 12/13/1	Tue 12/20/11
13	Submit Project Overview	6 days	Wed 12/21/1	Wed 12/28/11
14	System Development	100 days	Thu 12/29/1	Wed 5/16/12
15	System Deployment	14 days	Thu 5/17/12	Tue 6/5/12

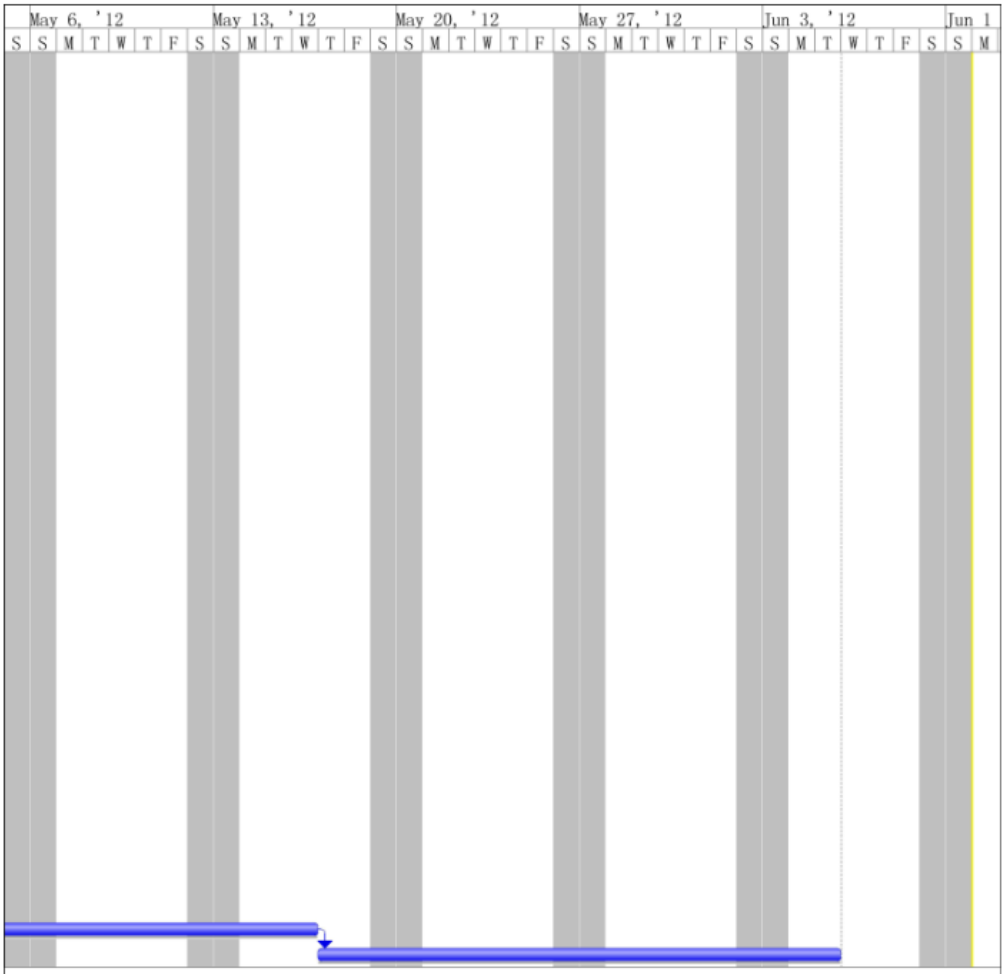
ID	Task Name
1	First Meeting with Supervisor
2	Prepare and Summit Chapter 1
3	Discuss About Title and Possible Alternative
4	Confirm Change of Title, Discuss About Correction
5	Discuss About Literature Review, Submit Literature Review I
6	Continue with Literature Review Part II
7	Submit Part II of Literature Review
8	Correction on Literature Review
9	Complete Literature Review
10	Proceed with Chapter 3 Methodology
11	Submit Correction of Chapter 2, Submit Chapter 3
12	Discuss About Project Overview
13	Submit Project Overview
14	System Development
15	System Deployment




, '11
Sep 25, '11
Oct 2, '11
Oct 9, '11
Oct 16, '11
Oct 23, '11
T W T F S S M T W T F S S M T W T F S S M T W T F S S M T W T F S S M T W T
Page 2


Oct 30, '11
Nov 6, '11
Nov 13, '11
Nov 20, '11
Nov 27, '11
Dec
F S S M T W T F S S M T W T F S S M T W T F S S M T W T F S S M T W T F S S
Page 3


4, '11
Dec 11, '11
Dec 18, '11
Dec 25, '11
Jan 1, '12
Jan 8, '12
M T W T F S S M T W T F S S M T W T F S S M T W T F S S M T W T F S S M T W
Page 4


May 6, '12
May 13, '12
May 20, '12
May 27, '12
Jun 3, '12
Jun 1
S S M T W T F S S M T W T F S S M T W T F S S M T W T F S S M T W T F S S M
Page 8