

ECOMMERCE USING SEARCH ENGINE OPTIMIZATION (ESEO)

YEAP JIA WEI

BACHELOR OF COMPUTER SCIENCE
(SOFTWARE ENGINEERING)
UNIVERSITI MALAYSIA PAHANG

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

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 memperbaiki 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 & CONCLUSION

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 web servers. But both the datacenters and web servers 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.

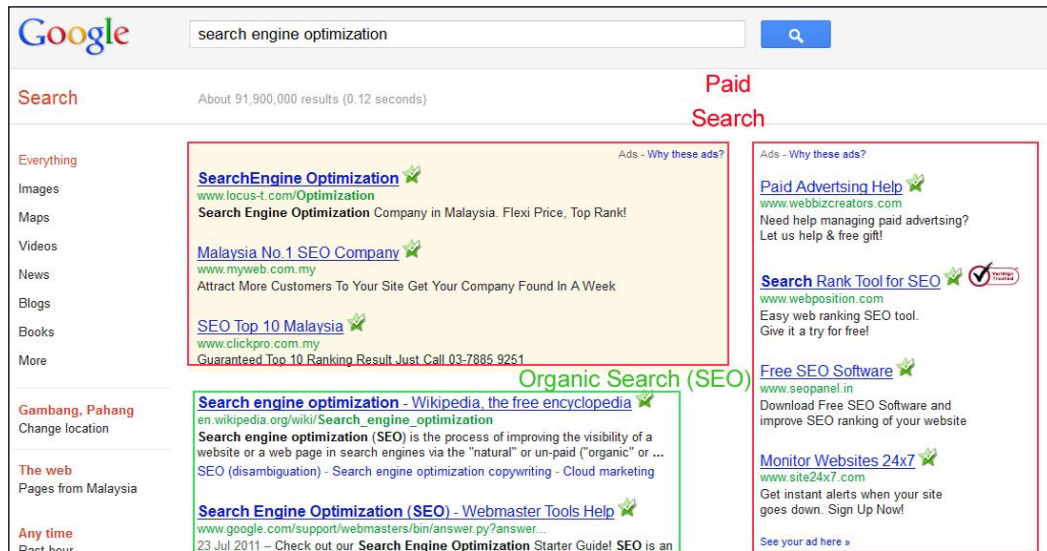


Figure 2.1: Different Sections of Search Engine Results Page [24]

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.

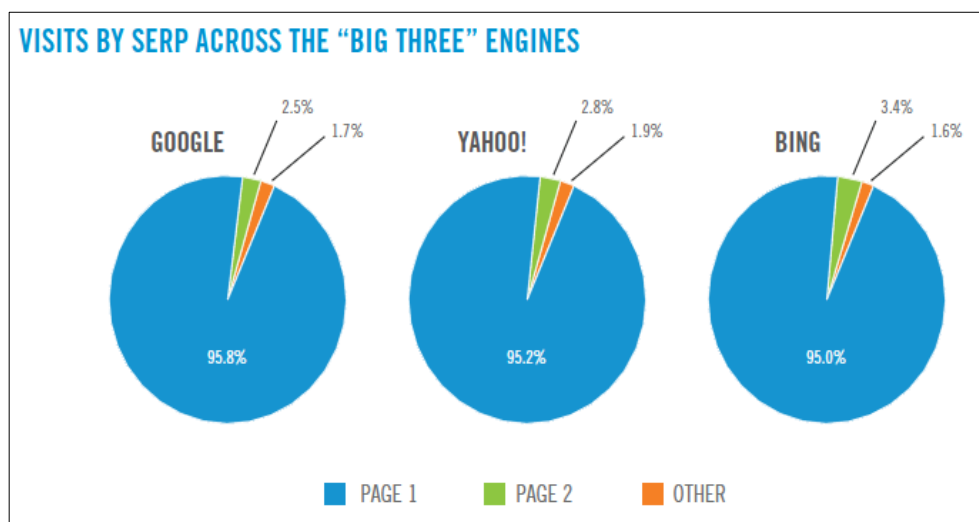


Figure 2.2: Visits from Three Major Search Engine Results Page [5]

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.

Locations and Languages

Locations ? In what geographical locations do you want your ads to appear?

All countries and territories

Malaysia

Let me choose...

[Show map](#) | [Send feedback](#)

Languages ? What languages do your customers speak?

English [Edit](#)

[Advanced location options](#)

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.

Bidding and budget

Bidding option ? Basic options | [Advanced options](#)

Manual bidding for clicks

You'll set your maximum CPC bids in the next step.

Automatic bidding to try to maximize clicks for your target budget

Budget ? MYR per day

Actual daily spend may vary. ?

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.

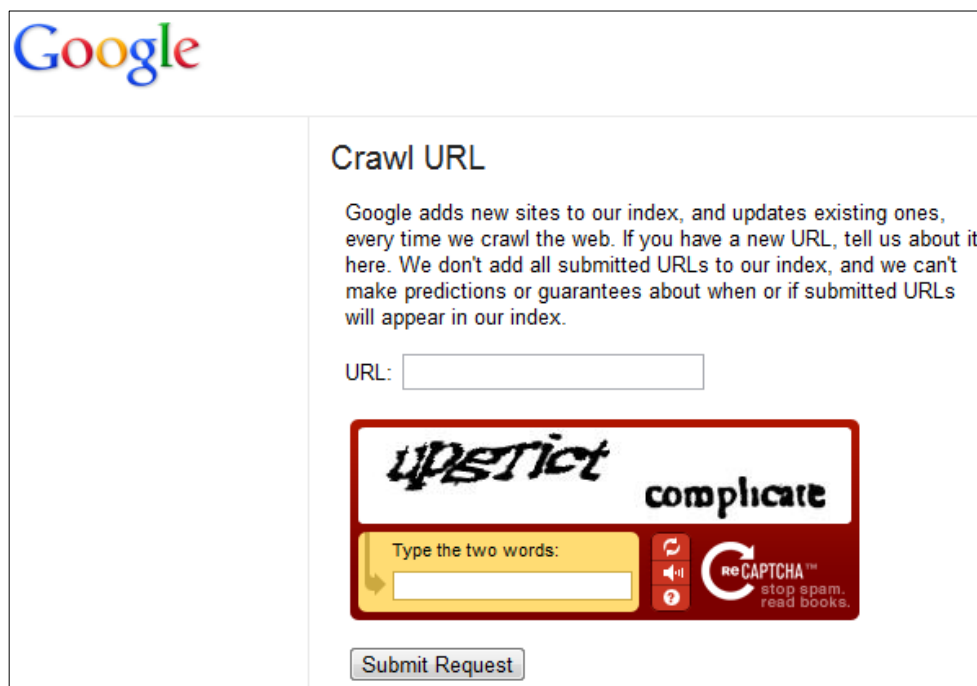


Figure 2.7: Google Webmaster Tool (Crawl URL) [29]

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.

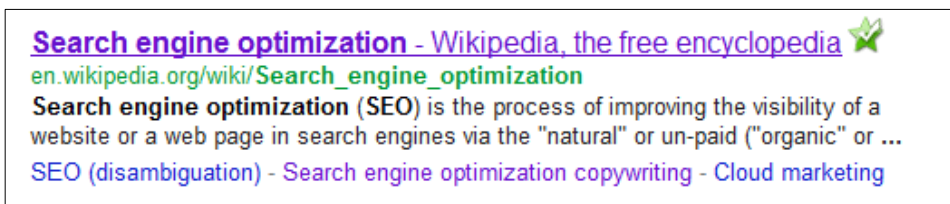


Figure 2.9: Title Tag Appears as Headline on Google SERP [24]

```
<html lang="en" dir="ltr" class="client-chrome client-chrome-15 client-webkit client-win client-js"
xhtml">
<head>
<title>Search engine optimization - Wikipedia, the free encyclopedia</title>
```

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.

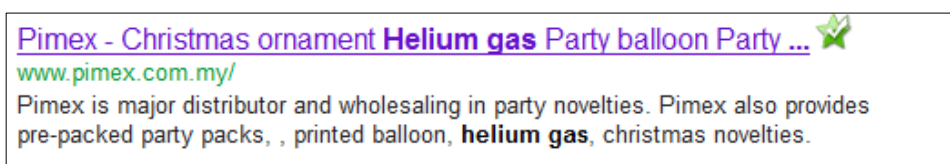


Figure 2.11: Description Meta Tag Appears as Snippet on Google SERP [27]

```
<meta content="Pimex is major distributor and wholesaling in party novelties. Pimex also provides
pre-packed party packs, , printed balloon, helium gas, christmas novelties. Aside of that, we also
welcome for custom made balloon" name="description">
```

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.