Master Thesis Computer Science Thesis no: MCS-2011-10 March, 2011



Exploring Search Engine Optimization (SEO) Techniques for Dynamic Websites

Wasfa Kanwal

School of Computing Blekinge Institute of Technology SE – 371 39 Karlskrona Sweden This thesis is submitted to the School of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Computer Science. The thesis is equivalent to 20 weeks of full time studies.

Contact Information:

Author:

Wasfa Kanwal

E-mail: wasfa.kanwal@yahoo.com

University advisor: Martin Boldt, PhD. School of Computing

School of Computing
Blekinge Institute of Technology
SE – 371 39 Karlskrona

Sweden

Internet : www.bth.se/com Phone : +46 455 38 50 00 Fax : +46 455 38 50 57

ABSTRACT

Context: With growing number of online businesses, Search Engine Optimization (SEO) has become vital to capitalize a business because SEO is key factor for marketing an online business. SEO is the process to optimize a website so that it ranks well on Search Engine Result Pages (SERPs). Dynamic websites are commonly used for e-commerce because they are easier to update and expand; however they are subjected to indexing related problems.

Objectives: This research aims to examine and address dynamic websites indexing related issues. To achieve aims and objectives of this research I intend to explore dynamic websites indexing considerations, investigate SEO tools to carry SEO campaign in three major search engines (Google, Yahoo and Bing), experiment SEO techniques, and determine to what extent dynamic websites can be made search engine friendly on these major search engines.

Methods: In this research, detailed literature survey is performed to evaluate existing knowledge for SEO for dynamic websites. Further empirical experiments are conducted to address dynamic websites indexing problems; and to evaluate SEO techniques used in empirical experiments.

Results: It is found that all major search engines, including Google, cannot fully index dynamic websites. I used some SEO techniques which I explored during this study to help dynamic webpage(s) get indexed in major search engines. The experiment results reflect the effectiveness of SEO techniques including URL encoding /friendly URLs on major search engines.

Conclusions: I conclude that, dynamic websites are subjected to indexing related problems and require additional SEO efforts to appear in SERPs. Not all SEO techniques are equally effective on all search engines to improve indexing of dynamic webpage(s). Each implemented SEO technique has different impression on major search engines (Google, Yahoo, Bing, Ask, and AOL). As, the encoded URLs technique is effective on all major search engines. However, Yahoo and Bing prefer friendly URLs over typical URLs with parameters. Therefore, presentation of dynamic URL could be quite paying if it is needed to index dynamic website on search engines other than Google.

Keywords: Search Engine Optimization, Dynamic Websites, Search Engine Friendly

CONTENTS

ABSTRACTII			
T	ABLE (OF CONTENTS	111
L	IST OF	TABLES	VI
L	IST OF	BLE OF CONTENTS 111 ST OF TABLES VI INTRODUCTION 1 .1 BACKGROUND 2 1.1.1 Search Engines 2 1.1.2 Paid Result vs. Organic Results 2 1.1.3 Static Website vs. Dynamic Website 3 1.1.4 Website Indexing vs. Website Ranking 4 1.1.5 Visible Web vs. Invisible Web 4 1.1.6 Search Engine Optimization (SEO) 4 .2 RELATED WORK 5 PROBLEM DEFINITION 7 2.1 PROBLEM OUTLINE 7 2.2 OBJECTIVES AND GOALS 7	
1	INT	RODUCTION	1
	1.1	RACKGPOLIND	2
		e e e e e e e e e e e e e e e e e e e	
	1.1.4		
	1.1.5		
	1.2	RELATED WORK	5
2	PRO	BLEM DEFINITION	7
	2.1	PROBLEM OUTLINE	7
	2.2	OBJECTIVES AND GOALS	7
	2.3	CONSTRAINTS	8
3	RES	EARCH APPROACH	9
	3.1	MOTIVATION AND RESEARCH QUESTIONS	9
	3.2	RESEARCH METHODS	9
	3.3	HYPOTHESES FORMULATION	10
4	LIT	ERATURE SURVEY	11
	4.1	SEARCH ENGINE OPTIMIZATION (SEO) TECHNIQUES	12
	4.2	ON-PAGE SEO TECHNIQUES	
	4.2.1		
	4.2.2		
	4.2.3	6	
	4.2.4	U	
	4.2.5	O .	
	4.2.6	3	
	4.2.7 4.2.8		
	4.2.9		
	4.2.1		
	4.2.1	* * *	
	4.2.1		
	4.3	OFF-PAGE SEO TECHNIQUES	
	4.3.1		
	4.3.2		
	4.3.3	3	
	4.3.4		
	4.4 4.4.1	IMPORTANT DESIGN CONSIDERATIONS FOR SEARCH ENGINE FRIENDLY WEBSITES	
	4.4.1 4.4.2	-1 3	
	4.4.2 4.4.3	•	
	4.5	ENCODING URLs FOR DYNAMIC WEBSITES	
	4.5.1		
	152	O Company of the comp	21

- 3	
OVERVIEW OF THREE MAJOR SEARCH ENGINES AND THEIR INDEXING CONSIDERATI	ONS FOR
IC WEBSITES	3
Overview of Google (Search Engine)	3.
, ,	
<i>7</i> 1	
Selection of Variables	50
Selection of Subjects	5
· ·	
Validity Evaluation	5
EXPERIMENT OPERATION	5
Instrumentation	5 [,]
APPLIED SEO TECHNIQUES	5′
·	
1 1	
Friendly URLs	6
RESULTS ANALYSIS	6
Outliers	6
Efficiency	6
Evaluation of indexing results	6
· ·	
•	
Hypotheses Testing	
EMPIRICAL STUDY	
CLUSION AND FUTURE WORK	7
Conclusion	
FUTURE WORK	7
ALCEC	7 !
NCES	•••••
	Method for URL rewriting OVERVIEW OF THREE MAJOR SEARCH ENGINES AND THEIR INDEXING CONSIDERATI IC WEBSITES Overview of Google (Search Engine) Dynamic websites Indexing considerations of Google (search Engine) Dynamic websites Indexing considerations of Yahoo (Search Engine) Dynamic websites Indexing considerations of Yahoo (Search Engine) Dynamic websites Indexing considerations of Microsoft's Bing (Search Engine) Dynamic websites Indexing considerations of Microsoft's Bing (Search Engine) Dynamic websites Indexing considerations of Microsoft's Bing (Search Engine) Dynamic websites Indexing considerations of Microsoft's Bing (Search Engine) USEFUL SEARCH ENGINE'S TOOLS FOR SEO CAMPAIGN Google's Useful Tools for SEO Yahoo's Useful Tools for SEO Webrank Toolbar for Firefox ERIMENT GOAL DEFINITION EXPERIMENT PLANNING Hypothesis Selection of Variables Selection of Subjects Experiment Dexign Validity Evaluation Experiment Dexign Validity Evaluation EXPERIMENT OPERATION Instrumentation EXPERIMENT OPERATION MEASUREMENT PREFACE APPLIED SEO TECHNIQUES RESULTS AND INTERPRETATION MEASUREMENT PREFACE APPLIED SEO TECHNIQUES RESULTS ANALYSIS Outliers Efficiency Evaluation of indexing results Evaluation of friendly URL's results Normality Testing Shapiro-Wilks Normality Test Non-Parametric Test Mann-Whitney Test Hypotheses Testing CUSSION LITERATURE REVIEW EMPRICAL STUDY GCLUSION AND FUTURE WORK

LIST OF TABLES

Tables	Page
TABLE 4.1 Robot Meta Tag values and their functionality	10
TABLE 4.2 Specifications that can be provided within Robot Meta Tag	17
TABLE 6.1 Webpage(s) indexing summary for website1 (wasfabththesis.com)	45
TABLE 6.2 Comparison of experimental and control group for SEO techni website1 (wasfabththesis.com)	-
TABLE 6.3 Webpage(s) indexing summary for website2 (recipe-planner.com)	40
TABLE 6.4 Comparison of experimental and control group for friendly URLs	23
TABLE 6.5 Efficiency of SEO techniques applied on website1 (wasfabththesis.c	om)30
TABLE 6.6 Efficiency of SEO techniques applied on website 2 (recipe-planner.o	com)10

LIST OF FIGURES

Figures	Page
FIGURE 3.1 Research Methods	10
FIGURE 4.1 Literature Survey Process	12
FIGURE 4.2 Example webpage shows use of page title and its appearance	14
FIGURE 4.3 HTML code for example webpage that contains Title Tag an in a website and their appearance in Google search engine	_
FIGURE 4.4 Example webpage shows keyword usage in URL, Page Title, webpage contents	
FIGURE 4.5 Example of problematic linking structure/internal linking of a	ı website19
FIGURE 4.6 Example of location based bread-crumb of a website	20
FIGURE 4.7 Example of Anchor Text in a webpage	26
FIGURE 4.8 Example of one-way and two-way linking of websites	27
FIGURE 4.9 Example webpage that is using three frames to present webpage(s) contents to search engine against a single URL	
FIGURE 4.10 Explicit search shares of major search engines	33
FIGURE 4.11 Google Webmaster Tool screenshot.	43
FIGURE 4.12 Google Analytics Tool screenshot	44
FIGURE 4.13 Google AdWordsTool screenshot	45
FIGURE 4.14 Yahoo site explorer screenshot	46
FIGURE 4.15 YSlow for FireFox screenshot	46
FIGURE 4.16 Bing Webmaster Tool screenshot	47
FIGURE 4.17 Webrank Toolbar screenshot	48
FIGURE 5.1 Pretest on Yahoo: Sitemap submission with friendly URLs	79
FIGURE 5.2 Posttest on Yahoo: Sitemap submission with friendly URLs	80
FIGURE 5.3 Pretest on Bing: Sitemap submission with friendly URLs	80
FIGURE 5.4 Posttest on Bing: Sitemap submission with friendly URLs	80
FIGURE 5.5 Pretest on Google: Sitemap submission with friendly URLs	83
FIGURE 5.6 Posttest on Google: Sitemap submission with friendly URLs	84
FIGURE 5.7 Pretest on Ask: Sitemap submission with friendly URLs	84
FIGURE 5.8 Posttest on Ask: Sitemap submission with friendly URLs	85
FIGURE 5.9 Pretest on AOL: Sitemap submission with friendly URLs	85
FIGURE 5.10 Posttest on AOL: Sitemap submission with friendly URLs	86
FIGURE 6.1 Descriptive representation of indexing throughout experime step by step application of SEO techniques for website1 (wasfabththesis.com	
FIGURE 6.2 Descriptive representation of indexing throughout experime step by step application of SEO techniques for website2 (recipe-planner.com	
FIGURE 6.3 Performance chart of SEO techniques of website1 (wasfabthth	nesis.com).64
FIGURE 6.4 Performance chart of SEO techniques of website2 (recipe-plan	nner.com) 64

FIGURE 6.5 Total number of indexed and not indexed static and dynamic webpason website1 (wasfabththesis.com)	
FIGURE 6.6 Total number of indexed and not indexed dynamic webpage(s website2	•
FIGURE 6.7 Comparison of Friendly URLs between experimental and control gradient for website (recipe-planner.com)	_

Chapter 1

1 INTRODUCTION

Internet, the global source of information has become an essential part of our everyday life, and is commonly used for e-commerce and social networking. Millions of people use it for a variety of tasks including shopping, banking, gaming, dating, online booking and social networking. With Wireless Fidelity (Wi-Fi), internet is now accessible on most of the mobiles and handheld devices. Therefore, many companies have rationalized their business processes to make everyday business convenient for their customers by providing online services. Online statements, online forms, bill payments and account recharging are few examples of this transformation.

According to China Internet Network Information Centre (CNNIC) statistics, 76.3% of users prefer using internet to access information over any other source of information [1]. Nowadays, many businesses rely on online advertisement and e-commerce. Websites such as Google and Yahoo are earning huge revenues from online advertisements. With the growth of online businesses, there are millions of websites uploaded on Internet and their number is multiplying. As a result of this increase, businesses need to compete against other online competitors selling similar products and services for increasing their sales figures. This has introduced the concept online ads and optimizations of the website for search engines.

People trust search engines to find a reliable business during search process; therefore, a search engine is a key resource to boost online businesses. Online businesses pay thousands of dollars to make their websites search engine friendly. comScore Inc.[17] reported that about 15.7 billions of searches are performed every month which is approximately 6100 searches/sec. These figures are evidence of the integration of search engines with our daily lives. Search engines not only give a better contour to the internet but they have become biggest source of global information retrieval. With advancements in search engines, internet users do not bother to memorize website address/URL to extract information from a specific website; rather they specify explicit keywords in search engine's search area to obtain desired information or to find desired website. The search process in response explores the entire available network resources and provides user with most related information as search results [1]. Google, Yahoo, Microsoft's Bing, Ask and AOL are some of the commonly used search engines.

Search engines are software(s) working at backend of the search processes. They crawl and index websites and collect necessary information i.e. keywords and phrases from the websites. These keywords and phrases reflect what the whole website is about. Later on, this collected information is stored in the databases of search engines. The search results for any keyword or set of keywords to find information could be hundreds of pages. However, it is a common practice that users do not go beyond first couple of pages in search results [7].

Search engines use special programs called Crawlers or Spiders that crawl and index websites. The Crawlers continuously explore the internet and include new websites in search engine databases, indexes or re-index websites accordingly. Search engines rank websites on basis of their content's quality [9]; and display those websites up in search results.

According to statistics, search engines can index only 40% of the websites [10]. Remaining 60% of websites are invisible to the search engines. These invisible/hidden websites includes dynamic websites as well [1] [4] [5] [10]. To bring those invisible websites in the search results require extra work to make them visible to the search engines. For dynamic websites, this is often done by making use of Search Engine Optimization (SEO) techniques.

SEO is one of the most important and leading Search Engine Marketing (SEM) activities. SEM as an online marketing that increase websites visibility in search engines to promote them [7] [13]; therefore, web developers are motivated to optimize their websites to obtain high ranking and improve searchability in search engines and increase business revenues. Online business(s), particularly shopping carts, usually use dynamic websites because they

need to update website contents more frequently, to include new products and manage shopping cart data.

SEO techniques are intentions to bring a website among top search results for some specific keyword(s). In most cases, website optimization is considered as a two step process. In first step, a team of developers creates a website; and in the second step website is handed over to SEO experts for website optimization.

The main motivation behind my work is to increase understanding of web developers and programmers about SEO; so that websites are developed with SEO perspective in mind, this will ultimately reduce overhead of optimizing websites after development. With the knowledge of SEO, the visibility of dynamic website could be improved in search engines to a good extent [3] [13]. I also want to divulge few myths about dynamic websites such as friendly URLs. Furthermore, I will explore and implement some of the practices that are commonly used to optimize dynamic websites to make them search engine friendly on major search engines (Google, Yahoo and Microsoft's Bing). SEO is necessary to improve visibility of the website to improve the volume of traffic to a dynamic website by writing necessary code.

1.1 Background

1.1.1 Search Engines

A search engine is considered as a source to promote a website and its associated business over internet. Search engines explore websites' contents to gather information about a website. Therefore, there is a need to optimize a website to make it search engine friendly. This will help to bring a website in top search results. The search engine programs are called "crawlers", "robots" or "spiders". The search engines are classified into two general categories: Crawler-based search engines and Human-powered directories; both works in fundamentally different manner.

The crawler-based search engines typically works in three steps [8]. Firstly, They crawl through the website; secondly, they analyze the webpage information for a targeted URL or keywords, evaluate the correspondence between webpage and search criteria, then they write this information in specific format in its index database; Finally, they extract webpage(s) in response of search query, containing most relevant information, from index database. The final result is presented in the form of hyperlinks and precise summaries of corresponding websites [7].

Human-powered directories rely on human review, category listing, or indexing. The site owner submits a short explanation of the entire website to such directories. The website owner is not aware which part of the submitted contents will be displayed as description in the final search results. The websites with valuable and meaningful contents is preferred to be reviewed and stored in web directory such as Open Directory, Google Directory, Yahoo Directory and Looksmart. Nowadays, it is common to present both human-powered-listing and crawler-based results. These types of search engines are known as hybrid search engines.

1.1.2 Paid Result vs. Organic Results

As discussed in the section 1.1.1 that hybrid search engines like Google and Yahoo presents two types of search results against search keyword(s). These results are categorized as "Paid results" or "Pay Per Click" (PPC) which is powered by "human-powered-directories"; other results are named as "Organic" or "Natural" results.

Most of search engines present paid search results at the top or right side of the Search Engine Result Pages (SERPs). Some search engines have their own policy to place such results. For instance, Google Adwords is determined by "bidding model" where businesses do biding for a Cost Per Click (CPC) to have their ad appearance in paid search results [3].

Advertisers bid only for those keywords or phrases for which they want to have visibility of their website in search results (Paid Results).

On the other hand, Organic/Natural results deal with obtaining top placement to become more visible in the SERPs without paying to search engines. Such results are powered by crawler-based search engines. Unlike paid advertisements, web developers build up their websites to obtain top position in SERPs to become more visible in the Organic/Natural results and it is free of cost in general. Websites with unique, quality, up-to-date contents with more back links would have higher position in the Organic results. Long term return on Investment (ROI) for appearance in organic results is much more valuable than appearance in paid results, as searchers are more likely to click on organic results[3] [7]; therefore, developers tend to optimize their websites to increase their website's visibility in organic results in search engines. The creation of website for making it search engines friendly to acquire visibility in search engines and obtain higher place in SERPs of search engines is an essence of SEO.

1.1.3 Static Website vs. Dynamic Website

Generally, websites are classified into two categories: Static and Dynamic. A Static website is one that is written using HTML with some basic scripting languages such as JavaScript. Static webpage(s) are not drawn from a database; instead, each webpage is considered as a separate document. Conversely, dynamic websites are created on the fly according to preferences that users specify in a form or on the bases of value(s) that user selects from menus. In result of user selections, dynamic webpage is created by extracting data from the affiliated databases [6] [13]. Therefore, to modify contents of a dynamic webpage, it might only require updating its database records rather than changing contents on the webpage. Dynamic webpage(s) do not physically exist like HTML webpage(s).

Dynamic websites are created by using a variety of programming languages such as PHP, JAVA, ASP and ASP.NET with combination of HTML tag [6] [7]. Static websites are easier to develop and cheaper to host than dynamic websites. Static websites are preferred over dynamic websites when a website is small and its contents do not need to update frequently. Someone with basic knowledge of website development skills can easily modify the contents of static website. Therefore, small businesses prefer using static websites to get have web presence. On the other hand, dynamic websites are preferred for large businesses or when there is a need to change or upgrade contents of website frequently. Dynamic websites have some advantages over static websites. For instants, it is possible to update contents of dynamic website by a person with little or even no knowledge of web development and scripting languages [7] [13].

Sometimes we can judge whether webpage is static or dynamic. As a dynamic webpage have some special characters like "?" or "&" in their URLs.

URL of a static webpage may look like this: http://www.yourDomainName.com/products/car-bikes.htm

URL of a dynamic webpage may look similar to this: http://www.yourDomainName.com/products/ref=menu_4?ie=UTF8&node=8908

Unfortunately, many search engines are not programmed to handle dynamic URLs [3] [7] [9] [13-16]. Therefore, dynamic webpage(s)/dynamic URLs are not considered as search engine friendly [7] [13 - 16].

1.1.4 Website Indexing vs. Website Ranking

As discussed in section 1.1.1, crawler-based search engines use automated programs called "crawler". The crawler crawls and reads webpage(s) later stores the retrieved information in a summarized format in the central repository [3] [7] [9]. This is known as indexing. When a searcher enters a desired search term, usually a set of keywords or phrase, to find desired information through search engine; the search engine pulls results from indexes (search engine databases), in response of searcher's provided query or keyword phrase [13]. The purpose behind crawling and indexing is to optimize speed and performance in finding relevant information for a search query.

Search engines can rank indexed websites only. That's the reason that newly uploaded websites do not appear in search results. Crawlers visit new and already indexed websites periodically. The crawling period and revisit time depends upon search engine's algorithms [9] [13]. The search engines assign ranks to webpage(s). It is tough science to track how webpage(s) are ranked. Although, a high ranked website means that it has most relevant information according to the specific keywords. The website with most related information to search query in displayed at the top in search results [7]. Web developers try to optimize their websites for higher ranking in search engines because higher placement of website in search results ultimately brings more users' click or traffic on websites and promote website's associated businesses.

1.1.5 Visible Web vs. Invisible Web

"Visible Web" or "Surface Web" is generally a collection of static webpage(s) that are connected by hyperlinks and such kind of Web is retrievable by common search engines [1-5]. Conversely, huge collection of information that exits and accessible by via WWW but is not accessible through general purpose search engine is referred as "Invisible Web" [1] [3] [5].

Generally, dynamic webpage(s), video/audio clips, flash movies, and files/documents in non-standard formats, are considered as "invisible Web" and are not indexed by conventional search engines. As in case of dynamic websites, resulting webpage(s) are dynamically created by extracting data from the underlying databases in response of user's search query. Therefore, such Web poses two problems for search engines. First, search engines' crawlers cannot make the selections from menu to "construct" these dynamic webpage(s). Therefore, most dynamic webpage(s) are not indexed in search engines [4] [5] [13 -16]. Second, search engines' spiders are not programmed to follow or index dynamic URLs. One reason can be that dynamic URLs may cause spiders to catch in infinite loop/spider trap [3] [4] [5]. As a result, a large number of webpage(s) remain invisible to search engines. Apart from that search engines are evolving to understand dynamic websites.

1.1.6 Search Engine Optimization (SEO)

When we consider website development either dynamic or a static, there is a need to consider all important factors that can increase traffic to the website directly or through search engines. More the traffic, higher the website ranking would be and higher the sales figure would become. It is a common practice that developers use several ways to create eyecatching and mind-blowing effects on the websites to make them look appealing and attractive to the website's users, but owner of website cannot get anything out such website, if user is not able to see or find the website through search engines.

Search engines provide us a platform to present or sell products or services, and SEO techniques help to promote businesses through the search engines. At the same time, a search engine facilitates the end user to search what they are interested in buying. SEO is the procedure of improving visibility of a websites or a webpage(s) in search engines via

"natural"/"organic" (un-paid) search results. SEO is an art to customize contents of a website to make it search engine friendly.

Crawler can read HTML based webpage(s) without any problem; however, dynamic websites/webpage(s) are not always searchable in all search engines. Google and Yahoo considered as most prevalent search engines for searching and indexing the web. On the other hand, it is either inexistent or far from perfect to search and index dynamic webpage(s) [6] [7] [12-16]. This is developer's responsibility to make dynamic websites search engine friendly or searchable for search engines. SEO techniques for dynamic web applications require little extra programming knowledge about search engine's behavior. Most of today's websites tend to include dynamic contents because dynamic websites are easy to update and manage using Content Management Systems (CMS). E-commerce websites, blogs, and forums are based on CMS.

Google claims that it has made some progress to deal with dynamic websites or dynamic URLs [14]. On the other hand, about 34% searchers are relaying on search engines other than Google. Not all search engines are programmed to crawl and index dynamic URLs [6] [13-16]. Therefore, dynamic websites require some extra endeavor to optimize and make them search engine friendly. Online businesses can be promoted by optimizing websites for Yahoo, Microsoft's Bing, Ask and AOL. Common SEO techniques used to optimize static website might not enough to optimize dynamic websites.

There are several known SEO techniques to optimize dynamic websites for common search engines. One practice to optimize dynamic websites is to develop static webpage(s) equivalent to each dynamic webpage(s) and by keeping them on website [12]. URL rewriting is another approach to avoid dynamic /complex URLs related problem. The URLs are rewritten by removing parameters and special characters which are problem for the search engine's spiders to read and index them [11]. Submission of crawler webpage(s) is another useful approach to improve visibility of dynamic webpage(s) [10].

In this research work, I will discuss how one can develop search engine friendly website. The targeted websites are developed using PHP and are dynamic in nature. I will also discuss several SEO techniques that could be used to make dynamic websites search engine friendly for Google, Yahoo and Microsoft's Bing. Another aim of my research work is to improve developers' knowledge of SEO when building simple CMS driven website without strange looking URL which search engine encourage to crawl. Further, using empirical study I will try to show to what extent these SEO techniques are effective to promote dynamic websites on Google, Yahoo and Microsoft's Bing to get added revenues. The process of applying SEO techniques would needs to invest time and money, but it will yield added revenues for businesses in return. Additionally, I will explore available useful tools for SEO to make SEO processes efficient and effective.

1.2 Related Work

Now a day, Dynamic websites are commonly used in e-commerce; however, dynamic websites are subjected to indexing and ranking related problems. Google claimed in late 2008 that it has made some progress to index and crawl dynamic websites and can treat dynamic website as good as static but current research reflects that Google still have problem to index dynamic websites[3] [13]. A number of researches have been done to address this issue. Unfortunately, not any previous researches completely address the issue. Most of studies done until now cover only improvement of ranking problems on Google. Nevertheless, Yahoo and Bing are also prominent search engines but they are still not able to crawl and index dynamic websites [7] [15-16]. Both search engines are growing day by day and they are considerably competitor of Google.

E. Enge et al. [7] and J. L. Ledford et al. [13] stated in their research works that dynamic websites' URLs are not search engines friendly; because, many search engines are not programmed to handle dynamic URLs. Therefore, dynamic webpage(s)/dynamic URLs are not considered as search engine friendly. J. L. Ledford et al. [13] says that use of SEO

techniques can improve visibility of dynamic website; also SEO techniques could help to solve indexing problems in search engines to a good extent.

- C. Duda and G. Frey [12] presented a model of AJAX search to reflect indexing of AJAX application. The proposed demo presents the possible stages for AJAX search engine i.e. crawler, indexer and query processor. The research aimed to reflect the problems and challenges as well as solutions for indexing/ranking of AJAX based application.
- N. Nazar [9] has worked to explore SEO techniques for Web2.0 websites. The main focus of study was to improve ranking of Web2.0 websites. Though, his work is limited to explore problems related to CMS system which generate AJAX based dynamic website. Moreover, his work was mainly focused to enhance the capabilities of exiting CMS. His research was limited to improve ranking on Google.
- A. Pirkola [10] also worked to explore effectiveness of different search engines to index domain names from different countries. Her research was aimed to explore effectiveness of indexing on US based search engines i.e. Google, Live search, Virgilo, Voila and www.fi.
- G. Rogan [8] also worked to determine the effects of SEO methods for improvement of the ranking of websites. He identified some SEO methods by performing case studies on different websites to explore the effectiveness of indentified methods for better ranking on search engines. His research work was aimed to improve ranking of websites on Google search engine.
- J. Köhne [6] developed a model for resolving crawling/indexing issues of a specific CMS generated dynamic website. His research work identified some problems like parameters in URLs, keywords and site structures. He developed a model to resolve these issues to improve crawling and indexing of dynamic websites in Google.

Most recently, Dr. K. Baskaran and R. Vadivel [11] worked to implement SEO techniques on static as well as dynamic websites. Their focus of research work was to generate friendly/clean URLs for Model View Controller (MVC) web applications. Their focused of research was to explore the consequence of some SEO techniques for making cleaning dynamic URLs to make them keyword oriented. Moreover, they worked to explore implementation of URL rewriting and redirecting by using ASP.Net.

The above mentioned related works mainly focused to improve ranking of websites or by keeping some kind of dynamic contents of website in mind. However, I aim to explore the indexing related problem of dynamic websites and explore SEO techniques to improve indexing of dynamic websites. Furthermore, the focus of my study in not only limited to Google; but, I will also try to explore SEO techniques to improve indexing of dynamic websites in Yahoo and Bing.

2 PROBLEM DEFINITION

This chapter aiming to outlines the problem domain, aims, objectives, and constraints of this research work.

2.1 Problem Outline

Websites' visibility in search engine plays a major role in promoting online businesses; therefore, web developers are motivated to optimize websites to obtain high ranking and visibility in search engines to increase business revenues. Dynamic websites are not considered as search engines friendly for most of prominent search engines like Yahoo and Microsoft's Bing [3] [6] [14] [15] [16]. They are still not perfect to crawl and index dynamic webpage(s) [7]. Google and Yahoo are considered as most prevalent in searching web. Among other search engines, Google's crawler is recognized as rather precise at finding dynamic webpage(s), images, and other types of static content on the web. On the other hand, the contents of dynamic websites are stored in the databases and have no fixed addresses or URLs. Therefore, Most of search engines (including Google) are either inexistent or far from perfection for searching dynamic webpage(s) [12] [13].

Google claimed in late 2008 that it has made some progress to index and crawl dynamic websites and can treat dynamic website as good as static one [14]. Now the question is: how much progress have been made to cope with dynamic website indexing issues? Further, a comScore's statistics report says that 34% of people rely on search engines other than Google. Yahoo and Microsoft's Bing comes at 2nd and 3rd position for covering search area and searcher preferences [17]. These search engines are not able to index dynamic webpage perfectly and recommend to avoid dynamically generated URL [3] [12] [15] [16].

Thus problem can be defined as:

It is not yet known, that how much progress has been made by three major search engines (Google, Yahoo and Microsoft's Bing) for indexing dynamic websites? Although Google claims it has made good advancements in indexing dynamic websites and suggests not performing common known optimization techniques called URL rewriting. There is a need to know, is it possible to improve visibility of a dynamic website in Google by applying some SEO techniques to make dynamic websites search engine friendly? Also, that would SEO techniques affect visibility of dynamic website in Yahoo and Microsoft's Bing? Moreover, to what extent SEO techniques make dynamic websites search engine-friendly with regards to the three major search engines?

2.2 Objectives and Goals

The main aim of this study is to analyze and explore the SEO techniques to make dynamic websites search engine friendly on three major search engines (Google, Yahoo and Microsoft's Bing). For this purpose, I will be performing some experiments on dynamic website by applying different SEO techniques to make it search engine friendly on Google, Yahoo and Microsoft Bing. As discussed in section 2.1 that Google claims that it can deal with dynamic website to index and it does not have any problem with non-friendly/dynamic URL. Further, Google recommend avoiding conversion of dynamic URLs to friendly URLs. On the other hand, researchers continuously debates that search engines (including Google) are far from perfect to index dynamic URLs [7] [13]. Therefore, I will also perform an experiment to test behavior of major search engines on this controversial technique (i.e. conversion of dynamic URLs to search engine friendly URLs). In other words, I will

implement some SEO techniques to make dynamic website's URLs search engine friendly and will measure how the major search engines respond for indexing webpage(s) with friendly URLs and with dynamic URLs. I have to attain subsequent objectives to accomplish my aim:

- Investigate practices/SEO techniques for dynamic website to make them search engines friendly with regards to three major search engines.
- Explore considerations of three major search engines for dynamic websites indexing.
- Explore misconceptions about making dynamic websites search engine friendly with regards to the three major search engines.
- To investigate useful tools for SEO promoting SEO campaign on three major search engines.

2.3 Constraints

When a search engine responds to a search query, it can result in hundreds of pages with millions of results. However, it is common practice that searchers do not go beyond first two or three of pages in search results [3] [7] [9] [13]. As mentioned in chapter 1 (section 1.1.3), top search results are taken from the websites bearing relatively higher rank. Generally, aim of SEO is to improve website's visibility in search engine and obtain higher rank. However, my research work aim to target on the possible ways to improve indexing of dynamically generated webpage(s). Because In many cases, ranking takes at more than three months to get clear results of how successful the efforts are for improving website's ranking in search engine [7] [9] [13]. Due to time constraints, I will focus only those factors which can help improve indexing of dynamic websites. Crawler-based search engines crawl newly built and already indexed webpage(s) periodically; the frequency of visits time depends upon search engine's algorithm [13]. However, ranking of website can be improved steadily even after a site indexed by search engine spiders. I would like to mention here that I will consider indexing of a sample dynamic websites only on Google, Yahoo and Microsoft's Bing. But, I will also present indexing responses of other search engines i.e. Ask and AOL on my sample websites.

3 RESEARCH APPROACH

This chapter aims to describe the motivation of research work, research questions and implementation of research methods.

3.1 Motivation and Research Questions

It is believed that an appropriately indexed website has better chances to obtain better rank in search engines [7] [9] [13]. A beautiful looking website that cannot appear in search results will fail the efforts of developer(s), because a search engine is a mean to bring a website (business product) to the searcher (customer). A poorly optimized website would waste all efforts and money to promote it; therefore, it is a one of SEO facet to ensure a website will be indexed and ranked properly on most search engines [3] [7] [9] [13].

Based on my initial research about dynamic websites' indexing related issues, I found that there are many misconceptions about the indexing of dynamic URLs and how search engines interpret them. Now SEO is not the process of stuffing Meta tags with carefully chosen keywords, and providing concise description of the webpage; rather, it is a strategic and sophisticated methodology. I also found that there are many contradictory views about the need of friendly URLs. I found this area most interesting; so I decided to perform literature survey and empirical research method on this topic. Current research reflects that dynamic websites face more indexing problems than static websites on most search engines [3] [12-16]. Hence, I intend to address the following questions:

RQ 1: What are the state-of-the-art SEO techniques for dynamic websites, and how are these techniques implemented within the three major search engines?

RQ 2: To what extent can these state-of-the-art SEO techniques make dynamic websites search engine-friendly with regards to the three major search engines?

3.2 Research Methods

This research work is kind of descriptive study that involves analyzing and evaluating exiting knowledge and practices in the field of web development, designing and promotion/marketing. I will mainly focus in the area of development and designing.

In this research work, I will use two research methods one is literature survey [19] and other is empirical research method [18], to approach the two research questions addressed above [18].

RQ1 will approach through, detailed and comprehensive literature survey, aiming to reveal and explore the current-state-of the art for SEO techniques for dynamic websites. Further, I will explore three major search engines and their indexing considerations for dynamic websites. I will continue to perform literature survey to explore useful tools to optimize websites for three major search engines.

To approach RQ2 empirical experiments will be conducted to explore useful SEO practices/techniques to make dynamic websites search engines friendly. These practices will be practically applied to a PHP based websites that I will be using to do programming experiments to make dynamic webpage(s) search engine friendly. I will not be writing code from scratch to keep my focus on SEO techniques; instead, I will be customizing pre-written code for a dynamic website. The results will help me to evaluate the SEO techniques and to find best practices to make dynamic website search engine friendly on three major search engines. From the results of empirical experiments, it will be easy to determine the extent to

which a dynamic website can be optimized to make it search engine friendly. Empirical experiment will be conducted in methodological and logical manner on the based on statistical and empirical analysis of collected validated data.

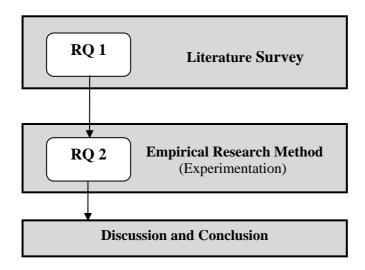


Figure 3.1 Research Methods

3.3 Hypotheses Formulation

I formulated the following hypotheses to answer the experimental based RQ 2 of my thesis.

Null Hypotheses:

Ho1: Dynamic websites do not require search engine optimization (SEO) since three major search engines are equipped to index them as it.

*H*₀₂: *SEO* techniques do not make significant difference in indexing dynamic websites on three major search engines.

Alternative Hypotheses:

*H*11: Dynamic websites require search engine optimization (SEO) for three major search engines to properly index them.

*H*₁₂: SEO techniques make significant difference in indexing dynamic websites on three major search engines.

Given hypotheses will allow me to make some conclusion about indexing dynamic websites on three major search engines. H_{01} counter H_{02} . H_{02} will be evaluated only when H_{01} be rejected.

During experiments, I will apply few SEO techniques in order to study targeted search engines' behaviors to index dynamic websites. This will include techniques that are common to both static and dynamic websites. If I will be able to prove alternative hypothesis (H_{II}) this can be proved that three major search engines, are not equipped to properly handle dynamic websites. So there is a need to make dynamic websites search friendly to make them visible in search results Further to determine to what extent it is possible to make a dynamic websites search engine friendly, I will examine search engines' response towards friendly URLs.

4 LITERATURE SURVEY

Literature survey consists of two main elements: literature search and literature review. These are essential parts of research process [19]. Literature Survey is mainly based on academic publications and relevant books. The topics of literature surveys are selected to support research work to gather relevant information regarding practical work done in relevant study area. Main purpose of literature survey is to familiarize researchers with main concepts, methods, and applications. This knowledge gives context and rationale to researcher's work.

In this chapter, literature review of relevant research work is presented. The Objective of this study is to review the exiting knowledge and empirical evidences regarding current state-of-the-art of SEO techniques for dynamic websites, and how these techniques are implemented for the three major search engines. To summarize, synthesize and critique literature material, I used literature survey methodology. In addition to that I used literature study for dynamic websites and SEO techniques. I particularly explored literature regarding SEO techniques that can improve dynamic website's visibility in major search engines, and other related concepts that used and referred in this research study were also considered.

The process of literature survey can be identified by splitting it into following steps:

- Identify main issues
- Select source of information
- Search/Refine searched material
- Summarize/Synthesize/Critique of search material

In this literature study, a thorough literature search and review is performed to entailed available resources to answer my research question. In first phase, I identified issue to approach literature search and review sources. My main focus was to find Books, research papers, articles, journals, conference proceedings, company white papers and reports and websites related to my research topic. To find related material, I approached different resources like ACM Digital Library, IEEE Xplore Digital Library, BTH Library, other online Libraries, and websites related to my research topic. I would like to mention here - Since search engines are playing a major role in my research work; therefore, I used official websites of Google, Yahoo and Bing to obtain most up to date information regarding their indexing consideration of dynamic website, their available SEO tools and reviewed many online articles from their official websites regarding my research topic. In the second phase of literature survey process, before utilizing the sources in literature review study, I summarized, synthesized and critiqued all of collected and available sources. In last phase, I put them these sources in my literature review study. Following figure is representing literature survey process design.

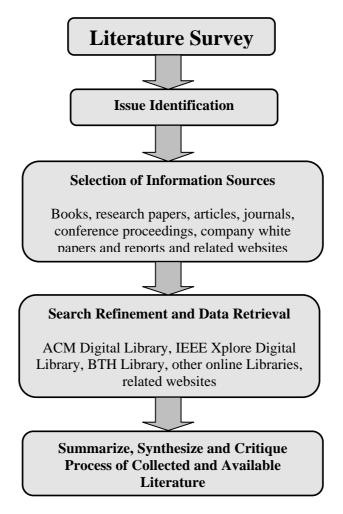


Figure 4.1 Literature Survey Process.

In this chapter, I intended to explore and emphasize the current state of the art SEO for dynamic websites, to explore three major search engines and their considerations for dynamic websites indexing; and eventually to investigate useful tools to optimize websites for these major search engines.

4.1 Search Engine Optimization (SEO) Techniques

The search engines are kind of platform for virtual marketplace for the potential buyers and sellers. SEO is all about how to optimize websites with the aim to make seller's website more visible in search engines to grab the searcher's/audience's attention to website and boost traffic. Indeed, SEO is very broad term that it is almost impossible to explain, if someone tries to understand it at once. Overall goal of SEO is to bring website on top to search results (organic results) [3] [7-9] [13] [20-23]. In businesses, when website is build, the main goal is to divert more and more traffic to website through targeted search engines.

SEO techniques/strategies are thought as some key tactics and ethical steps which should be considered while developing websites [3] [6-9] [13]. Although, the goal to bring a website among top ranked websites is not a dream that will come true over night. However, SEO is a long term process which continues with life of websites. The SEO techniques make it possible to tell the targeted search engines that what your website is about; finally it gives you a nice way to get targeted user/customer to visit your website through search results in targeted search engine.

Nowadays, companies develop websites and afterward they hire SEO experts to optimize their websites to make them search engine friendly. There is a misconception that SEO techniques are some magical spills that only possible to do by "SEO experts". Unlike this thought, a web developer who may not be a SEO expert, but she can take care of some basic points during development of website to optimize her website to make them search engine friendly.

At broad level, SEO techniques are recognized as containing two main components, On-page SEO techniques and Off-page SEO techniques [9] [13] [21] [23-26]. To be honest, both of them are essential and need to consider side by side for achieving goal of SEO for static websites as well as for dynamic websites. In my study, I am aiming to focus only dynamic websites crawling/Indexing issues in major search engines (Google, Yahoo, and Microsoft's Bing). There are several tools provided by search engines [8] [9] [13] [20-23] [26] to analyze the crawling/Indexing and ranking of webpage(s). Search engines crawls the websites and extract information according to their own criteria; later, this information is summarized to save in search engine's databases; and finally this saved information is presented as search result of websites in SERPs in response of searched query. Therefore, making a website to aid appropriate crawling/Indexing could be a positive step towards getting better ranking.

In the subsections below, I will describe some of useful On-page and Off-page SEO techniques which should be considered essentially while optimizing dynamic websites; as discussed in chapter 1 (section 1.1.3) that dynamic websites are not considered as search engine friendly like static websites for most of the search engines [6-16].

4.2 On-page SEO Techniques

On-page SEO techniques are employed on webpage(s) to optimize them to increase their worth in specific search engines [9] [13] [21-23]. In other words, on-page SEO techniques are used to optimize factors that are related to contents of each webpage (what the users/searchers see on webpage(s)/websites) and structure of website (what search engines crawlers see on webpage(s)/website) [23-26]. These techniques mainly comprise page title, header tags, Meta tags, target keywords, keywords density, ALT tags, content placement, breadcrumb trail, URL structure and size, internal linking of webpage(s), site update Frequency; last but not least, sitemaps and robot.txt files. These factors are heart of on-page SEO techniques to make website friendly for both website's users and search engines [13] [23-26]. As these techniques are important for both website's users and search engines point of view; therefore, these techniques need to be implemented with a good care. These techniques put in the picture the theme and contents of targeted website. In the following subsections I will provide a brief introduction to the factors which should be considered while developing dynamic websites.

4.2.1 Page Title/Title Tag

Page title tag is one of the most significant tags in On-page SEO because it informs both search engines and website's users about contents of particular webpage. The title tag is represented as <title> and it is basically a HTML code in the <head> section. It is important because it is used to create a string of text that appears in the top bar of Web browser. Also, search engines display page title as a headline- with hyperlink to enter your website- in search engine results. The "page title" (title tag) is essential and critical factor due to another reason because almost each search engine ranking algorithms consider title of webpage while crawling/indexing and display title in search result as well [7] [9] [23]. During search engine's crawling process "page title" is a beginning point of crawlers. Moreover, searcher clicks on search result in search engine's SERP if she finds headlines/"Page Title" relevant to their search query. Title Tag of webpage considered as major factor in on-page SEO because of the following prominent reasons [9] [20-26]:

- Search engine's ranking algorithm expects the contents of webpage related to the title of webpage [7] [9].
- In SERPs, search engines display page title as a heading/headline in response of search query. The page title of webpage is displayed as text link to the website.
- Title of page is displayed in top bar of browser window as name of the page being viewed by a user; thus it has an important navigational usability for users and browser.

World Wide Web Consortium (W3C) recommends that length of page title should not be more than 64 characters (including spaces) because most of browsers and search engines truncate length of title of webpage to make it consistent to display in SERP[13] [26].

The search engine spiders take contents of <Title> to determine that what the contents of specific webpage are expected to deliver to the end user; therefore, it is always recommended to use keywords in title (In the beginning of title string). Use of separators like "|","-", and"." is better if title is combination of more than one keyword phrases. Although, these separators do not work as identifiers for the search engines' spiders, and does not bring SEO benefits but is increase readability of title and encourage users to click on search result which have readable and related title.

Use of apostrophe, comma and other special character should be avoided; if there is a need to use them it is better to use HTML code of the character to be used in the title. Not all search engines recognize apostrophe in the same way; it is found that Ask has problem searching webpage(s) against search keyword containing apostrophe in it.

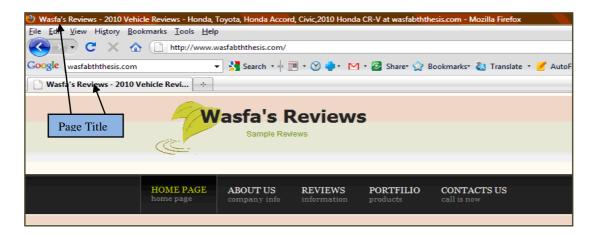


Figure 4.2 Example webpage shows use of page title and its appearance.

4.2.2 Meta Tags

Like <Title>, Meta tags are also placed in header section of a page i.e. between <head> tags of HTML code. Some of Meta tags are essential for having the website properly listing/ indexing in a search engines [9] [13] [20-23] [26]. The commonly used Meta tags are as follows: abstract Meta tag, keyword Meta tag, description Meta tags, expiry Meta Tag, distribution Meta tag, copyright Meta tag, robot Meta tag, language Meta tag. However, not all but some of these tags need specific attention because most of search engines consider them for indexing and ranking of website [20] [21] [25] [26]. The subsections will be provided with a few important Meta tags with some detail.

Keyword Meta Tag

Keyword Meta tag is a very important tag used by search engines to find a page for a searcher. It contains a series of important keywords specific webpage which ultimately

reflect/represent the contents of webpage. Those search engine which support Meta tags consider this tag for indexing of website [9] [11] [20].

Syntax of keyword Meta tag is as below:

```
<Meta name= "Keywords" contents="first_keyword, second_keyword,
nth_keyword"/>
```

Although, Google does not pay much attention to this Meta tag for ranking of webpage, however, search engines (including Google) consider to Mata tag for indexing (semantic indexing) [7] [23-26].

Descriptions Meta Tag

Description Meta tag is used to describe webpage(s) contents in short format but precisely containing keywords related to webpage. Some search engines use contents of description Meta tag as it is for indexing /listing. This tag plays a very significant role to improve Click Trough Rate (CTR) of website. Google Webmaster tool [34] provides some useful tips and cautions about using description Meta tag. Most of the search engines use the description Meta tag for getting insight of webpage [7] [13] [20]. Later, this snippet is used to display in search results. Therefore, this factor ultimately affects CTR of website.

Syntax of Description Meta tag is as follows:

```
<meta name="description" content="Description of webpage."/>
```

Some human-directories (search engines) use this description for listing of website in their indices. Therefore, this Meta tag important for both searcher and search engines because it gives a hint about website contents.

Robot Meta Tag

The robot Meta tag is specifically used to define rules for search engines regarding how to treat your webpage. The actual purpose of this tag is to guide and control crawlers for crawling and indexing webpage(s). The specified rules under this tag are applied to all search engines [7] [13] [20-23]. The multiple specifications can be provided under this tag. It may contain values like:

Table 4.1 Robot Meta Tag values and their functionality.

Values of Robot Meta Tag	Functions
Noindex	Do not index webpage.
index(default)	Index this webpage.
follow(default)	Follow hyperlinks on this webpage.
Nofollow	Do not follow hyperlink on this webpage.
Noodp	Do not use text from ODP (a.k.a. dmoz.org) to generate a title
	or snippet for this webpage.
Noarchive	Do not present "Cached" link for webpage in search results.
unavialable_after:[date]	Eliminate webpage from search results after specified time.

If this tag is not used then search engines by default consider crawling and indexing of webpage(s). This tag is very useful when it is needed to restrict search engine(s) for crawling non HTML files like (image, PDF files and other kind of doc files).

Sometimes website needs to have more than one version of same contents (.html, .pdf, .doc and print friendly version) which pose duplicate contents issue. Though, search engines are smart enough to find duplicate contents and index only one version but it consumes crawling time of a website. Fortunately, "nofollow" and "noindex" can be used to resolve this problem by restricting crawlers to crawl different versions of same contents. In saved crawling time, other important webpage(s) can be served; which ultimately improved indexing of website [7] [13] [22].

Syntax of robot Meta tag is as follow:

```
<META NAME="Robots" CONTENT="INDEX,FOLLOW">
```

Distribution Meta Tag

Distribution Meta tag is used to provide specification about the distribution of webpage(s) contents [20] [22]. In simple words, this tag specifies that in which areas/regions the website contents should be available through search engines. The contents of website would be available only according to the specified value under this tag. The values of this tag are as follows [22].

Table 4.2 Specifications that can be provided within Robot Meta Tag.

Values of Robot Meta Tag	Functions
Global	Available to the entire web
Local	Available to regional sites
IU	For internal use/Unavailable for public distribution

Syntax of distribution Meta tag is:

```
<META NAME="distribution" CONTENT="Global">
```

Expiry Meta Tag

Expiry Meta tag is very important where it is needed to specify that when webpage(s) is needed to be remove from search engine's indices/directory. This tag is very useful for frequently updated sites (i.e. news sites) to refresh search engines indices. This tag allows websites to get more places in search engine's databases for indexing new/updated page contents [13] [20] [23]. This can be set with specific date. For instance, if this tag is provided with "30 September 2010" then search engines would remove the webpage from search engines' indices on provided date.

Syntax of Expiry Meta tag is as below:

<meta name="expiry" content="never"/>

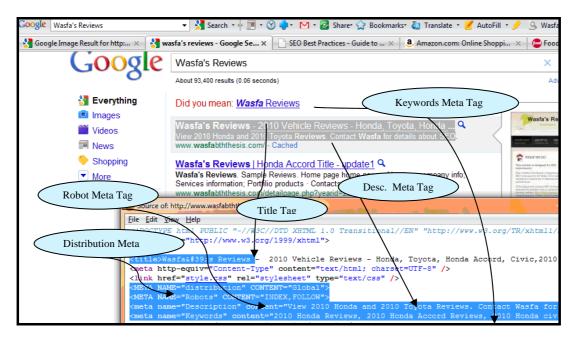


Figure 4.3 HTML code for example webpage that contains Title Tag and Meta Tags in a website and their appearance in Google search engine.

4.2.3 Targeted Keyword

The searcher enters keywords/keyword phrase in search engine's search area to obtain desired information for targeted keyword. Therefore, selection and placement of keywords is an essential element of SEO campaign. The discovery and decision about keywords should be taken even before the selection of domain name; since contents, title and URL of website need to have sufficient keywords [7] [13] [23-26]. However, an already exiting website can also be optimized by investing some energy and time for discovery of keywords. The discovery of keywords and their appropriate use in website contents leads towards better indexing and later for better ranking [7] [9] [13].

Developers make a common mistake that they try to rank their website for single word instead of a chain of keywords. This leads their website towards a continuous and lose of traffic (i.e. lose of 80% traffic) in search engines; because only 20% of searchers look for single word (search term). Whereas, 33%, 26% and 21% searcher search for two, three and four set of keywords- respectively- through search engines [23].

A common question arise in mind that what should be the *keyword density* of a webpage? Where *keyword density* means number of times the specific keyword appears in webpage contents. In simple words, it is ratio of keyword in webpage. The excessive use of keywords in a website/webpage sometimes becomes problematic; because, search engines would consider it as *keyword stuffing* [7] [13] [22-26]. This might cause that such website would not be indexed by search engines and in worst case website would be banned (search engines remove such website from their indices) [7] [9] [13] [25].

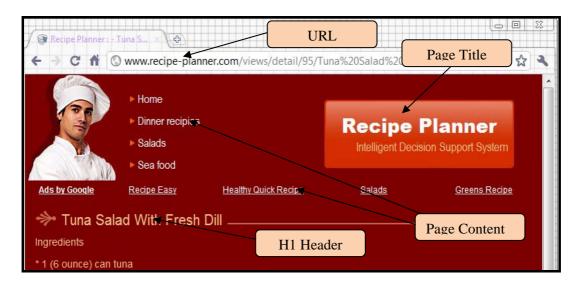


Figure 4.4 Example webpage shows keyword usage in URL, Page Title, Header and webpage contents.

4.2.4 Header Tags

Header tags are another important element of on-page SEO strategies. Those tags also exit inside HTML code just before
body> tag. HTML supports 6 level of heading [7] [13] [20]. Cascading Style Sheets (CSS) can also be use to handle these tags in a systematic way. When search engines spiders examine webpage, they also consider text/contents under this tag for including into the indices. Therefore, use of keywords in header tag is important for crawling point of view. It allows you to provide important keywords in an appropriate way. All search engines consider this tag; so there is no reason to avoid them [7] [13]. The figure 4.4 is presented in section 4.2.3 is an example of H1 Tag.

Syntax of heading tag (First level heading) looks like this:

<h1> most important Heading/set of keywords </h1>

4.2.5 ALT Tag

The use of images in websites is believed as more descriptive and eye catching than description in textual form. Though, user friendly contents of website might not be friendly for search engines; as search engines cannot access and understand every type of contents [7] [6] [9] [20-26]. Images are one of these troubling contents for search engines. It is always needed to make it sure that contents of website are also accessible and crawlable by search engines. It sounds simple to do this; but in fact it is not such easy to implement. As images are virtually invisible to search engines; although, ALT tags are helpful for making images visible to search engines [7] [13] [20] [21] [25]. An ALT tag is abbreviated name of "alternative tag". The ALT tags are used to provide textual description of image. This textual description tells the search engines that what is image about. Moreover, this tag is helpful when user try to access your website on that browser which does not support images, then this textual information about picture (under ATL tag) is presented to user as an alternative of image. Another advantage of these tags is that it makes images searchable to search engines. Google's image search feature is also based on this tag [22] [23].

Syntax of ALT tag is as follows:

4.2.6 Internal Linking

Search engines follow links on webpage(s) to discover other webpage(s) of the website [7] [13] [20] [23]. For this reason, developers should pay a good attention to build internal link structure of website. Some developers do a common mistake of hiding navigation or by making confusing navigations which makes website difficult to crawl and index.

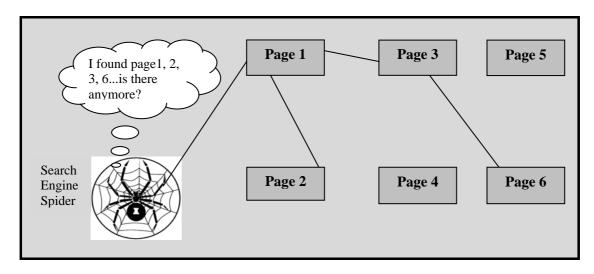


Figure 4.5 Example of problematic linking structure/internal linking of a website.

This example website's linking structure does not connect all webpage(s) of websites properly. It shows that "page 4" and "page 5" are not connected to any other webpage. Therefore, search engine's spider has no way to reach those webpage(s). Bear in mind it could cause search engine spiders to leave website without indexing disconnected webpage(s). In essence, it is very important to take care that each desired webpage(s) of a website are connected with proper navigations and reachable for search engines [13] [22] [25].

4.2.7 Content Placement

In SEO campaign, clear visibility and access to the desired contents on each webpage is not only desired for website's users. Nevertheless, it is necessary to provide a good presentation of important contents of webpage to search engine. It would help search engines to crawl and index websites more effectively [7] [9].

Sometimes, important contents of website which we need to get indexed are hided from spiders; because the desired contents are placed below in the page that might not be included in search engine indices. Web developers make a similar mistake that they place important contents in way more and more accessible for website users; though, a user friendly content placement might not be give you fruitful result of indexing. The placement of navigations to other webpage(s) in the beginning of webpage may cause spiders to switch to the next page and this cause spider to miss useful contents of current page [7] [13]. Therefore, it could be tricky to place contented on webpage in friendly way for both website's users and spiders. The placement of contents that are not search engines friendly like JavaScript should not place in the beginning of code [7] [20].

4.2.8 Bread-crumb Trail

Bread-crumb trail is a text based navigational approach. It is very helpful for both website's users and spiders for returning back to the previous webpage on the navigational path of website [13] [23]. It allows website's users to know where they are in the website which

make easier to travel through the website. Nevertheless, it makes easier for crawlers to examine website completely. It leads crawlers to crawl each webpage of website that is needed to include in SERP [9] [23]. It is also obliging to resolve duplicate webpage/contents issue in websites. Unfortunately, it is not ignored in most of website.

Typically bread-crumb trail is written in this way:

Index Page » Page1 » Sub Page 1 » Sub Page 2 » Sub Page 3

Generally Bread-crumb trail are divided into three different types [23]:

- **Location-based bread-crumb**: It informs that where webpage is located in hierarchy of website.
- **Path-based bread-crumb:** It informs that which path has been taken to arrive at the current webpage.
- **Attribute-based bread-crumb:** It delivers information about the category of current webpage.

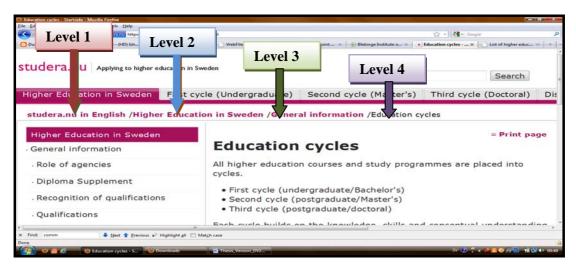


Figure 4.6 Example of location based bread-crumb of a website.

4.2.9 URL Structure and Size

Static and keyword targeted URLs are considered as best for website's users and search engine spiders [7] [13]; because, readable URLs reflect insight of a website. Readable URLs serve like a "name plate" of house which identifies the residence of the house.

Difficult to understand URLs are often called as dirty URLs [6] [11] because they are composed of special characters that are irrelevant to searchers/users [7] [13] [23].

URL of a static webpage could look like this:

http://www.Domain.com/products/car-Bikes.htm

URL of dynamic webpage might look similar to this:

http://www.domain.com/product/ref=sa_menu_lapnet4?ie=UTF8&node=5688

Unfortunately, most of the search engines' crawlers cannot crawl dynamic looking URL that contains special characters (like &, %,?) or sometime avoid to crawl them [13-16] [20-26]. Therefore, website URLs should be readable for users and structure should be well organized by keeping it understandable for spiders. Such static and readable URLs make most of search engine crawlers easy to follow and navigate website.

A complex and illogical URL structure compels users as well as spiders looking for something to struggle to discover [13].

Dirty looking URLs have some troubling aspects, like [7]:

- Long URLs containing punctuations in them are difficult to type.
- Long and Complex URLs are difficult to remember; as, such URLs do not provide any hint that what the target source contains or what function will be performed. Therefore, these URLs do not promote usability.
- Dynamic looking URLs might have security risks. Such URLs have query string follows "?" (Question mark). These types of URLs are often modified by hackers for attaching the web applications. The files extensions like .pl, .asp, and .jsp etc also give away some important information regarding implementation of a dynamic website that may be hacked by hackers.
- Dirty URLs can cause spiders to crash. Some web developers intentionally or
 unintentionally make an infinite number of requests that can catch crawlers in an
 infinite loop. This reason can cause crawler stuck on checking same webpage many
 times which is actually the same webpage with different URLs. Such websites are
 referred as spider traps. Therefore, some search engines avoid crawling dynamic
 website URLs.

Therefore, it is recommended by major search engines to avoid complex and very long URL [13-16] [22].

4.2.10 Site Update Frequency

Often, web developers misunderstand that On-page SEO techniques need to implement only once and later there is nothing to think about that again but this is not reality. There are many factors which need to be considered for maintenance of website contents even after website is index or got a top rank [7].

Most of web developers like to get a deeper and frequent crawl of their websites. Search engines like crawling unique content [7] [13] [23] [26]. Therefore, it is fruitful to update website regularly and frequently by adding new webpage(s) and unique contents; because, this is an obvious way to attract search engines to come back for crawling your websites and re-index them.

4.2.11 Page Compression

Today, user friendly applications became more and more demanding. A user-friendly webpage(s) sometimes need to add a lot of images it results heavy webpage(s). Therefore, it has become more challenging for developers to cope with issues like large webpage size and load time. This factor is important for both website's users and spider's point of view [20-23].

According to eMarketer research [9] [23], 16% users leave the webpage if it takes longer than 10 seconds to open. It can even loose large amount of traffic. Moreover, webpage that take longer time to load might not be fully cached by crawlers [7] [13] [22] [23]. In other words, search engines does not like to revisit and index your website if its response time is not good and it take much time to load webpage(s). Major search engines like Google, Yahoo and Bing have are some limits from for file size.

Fortunately, page compression tools can be used to make webpage(s) size according to desired size. Recently, many tools are available to compress webpage(s) size in effective way. Compression of images, cleaning up HTML and CSS can be implemented to reduce webpage size [20] [22].

Apache allows using GZip compression for PHP files; it allows compressing .css and .js files with minimum code. It is possible to apply do GZip pages with few lines of code in .htaccess file and php files itself.

Using GZip Compression on PHP page:

Update .htaccess and add following lines to it:

```
# ------
# GZip for compression
# ------
php_flag zlib.output_compression On
```

Include following php code lines in the beginning of the page:

```
<?
if (substr_count($_SERVER['HTTP_ACCEPT_ENCODING'],
  'gzip')) ob_start("ob_gzhandler"); else ob_start();
?>
```

Applying GZip Compression on CSS and JS files:

It is possible to apply GZip compression on CSS and JS files using few lines of code. If Apache supports GZip compression then few lines of code can decrease the size of page by 80%. When using GZip compression, page inflation is performed by the browser after page is loaded; therefore, it is necessary the browser supports the GZip compression (New browsers are equipped with this utility). Add following code line at the beginning of CSS/JS file and save CSS/JS file as PHP file for example CSSFileName.css.php (for CSS file) and JSFileName.js.php (for JS file).

CSS:

```
<? php header("Content-type: text/css"); ?>
```

JS:

```
<?php header("Content-type: text/js"); ?>
```

Now these files will be included in the php files header in a different format, as shown below:

CSS:

```
<link href="path/CSSFileName.css.php" rel="stylesheet"
type="text/css" />
```

JS:

```
<link href="path/JSFileName.js.php" type=" text/javascript" />
```

There are some online tools that allow determining webpage load time and also determine if a page has any compression performed on it such as dnsqueries.com/en/check_http_gzip.php and gidnetwork.com/tools/gzip-test.php.

4.2.12 Search Engine Essential Files

There are two essential files in SEO campaign: sitemap and robot.txt. These files are not only important for search engine spiders but also play a significant role for website's users to

access website contents [7] [13] [22]. These files play vital role to direct search engines on websites. These files become more significant for large websites and dynamic websites. The section below is presented with some brief description of search engine essential files.

Sitemaps

Generally, there are two types of sitemaps [22]. They are known as HTML sitemaps and sitemap.xml [27]. Sitemaps become more necessary if websites are large and dynamic. Search engines discover webpage(s) from other already indexed websites linking to your website; through internal linking and navigations of website. However, sitemaps also inform the search engines' crawler about the webpage(s) which might not be discovered directly [7] [23]. Previously, sitemaps were used only for website's users to provide navigation to the important webpage(s) of website. Today, sitemaps are used both for crawlers and website's users to inform them about important webpage(s) of website in the form of list of webpage(s) on website. The section below is provided with a brief description of each of these sitemaps.

HTML sitemaps

Traditional sitemaps are known as HTML sitemaps. Previously, they were designed to facilitate website's users to make their access easy on important webpage(s) of website [7] [9]. These sitemaps are HTML listing of webpage(s)-often in section-with the intention to help the website's user to find desired information through them [23]. Traditionally, search engines also relied on text-based sitemap (HTML sitemaps) [7]. Usually, they are presented as bulleted, outlined linked with anchor text to each webpage of website (or some of important webpage(s) of website). Although, sitemaps do not promote ranking of a website but they allow search engines to find desired webpage(s) easily and effectively for crawling an indexing.

• sitemap.xml

Google launched "Sitemap 0.84 Protocol" in 2005 which would be using sitemap of XML format Google's XML sitemap was designed for search engine crawlers to help them in finding data more efficiently and faster. The XML sitemap was designed in response of larger and complexly structured website's indexing problems [22]. The Google's XML Sitemap in now updated to XML Sitemap Protocol 0.9 and it is jointly supported by www.sitemaps.org [22] [27]. Recently, Yahoo, Bing, Ask and AOL had also agreed to support the XML protocol. This shared recognition of XML sitemap Protocol gives advantage of creating single sitemap for all search engines (Google, Yahoo, Bing, Ask and AOL).

Sitemaps are obliging in following conditions [7] [9] [23]:

- If website is large and containing dynamic contents.
- o If website is new and does not have enough linking websites (external links) to it.
- o If webpage(s) of website are not linked properly to each other (internal linking) i.e. broken links.
- o If website has some webpage(s) that cannot be discovered by search engines crawlers directly. Especially, if website has some Flash or AJAX featuring webpage(s).

Recently it is possible to place a single line in robot.txt (robot.txt is discussed in next section) to inform all search engines about location of XML sitemap of website [23]. Example statement which is needed to write in robot.txt is shown below:

Sitemap: http://www.domainname.com/sitemap.xml

The XML sitemap must include following elements [27]:

- Start with <urlset> .
- o Close with </urlset>.
- Specify namespace between <urlset> and </urlset> tags (which are set of rule). Namespace is preceded with "xmlns" which shows that it is as XML namespace.
- <url> tag is included as entry for each URL (top level in webpage(s)) as a parent XML tag.
- <loc> tag is included as entry for child for each <url> parent tag.

However, remaining tags are optional. An example code for sitemap.xml is provided below [27]:

Search engines like Google and yahoo provide facility to generate XML sitemap automatically [7] [20-23]. There are also some 3rd parties' sitemap generation tools. However, Google tool provides some additional functionality like detects updates in website and monitor website traffic automatically.

robot.txt

Another essential file which is purely developed for search engines is known as robot.txt [22] [23]. All reputable search engines' crawlers honourably follow robot.txt file of websites. This file is used to restrict search engine spiders from crawling or indexing some non necessary webpage(s) to save crawling and indexing time. As every time Search engine spiders come for crawling website content and bring some budget in form of number of webpage(s) to crawl/crawling time. Therefore, robot.txt file can lead to better indexing because in some cases it could be a better idea to restrict search engine bots from crawling image directory of website, as it will be wastage of website bandwidth and crawling time [7] [13] [23]. Moreover, it works like a tool which can be used to restrict spiders from indexing those contents of website which you do not want to present publically in search results. As a developer, we can have little more control over search engine's robots that they are allowed to do on website. In order to use this valuable file, it is required to have access to the root of domain. File name should be the same as "robot.txt" (with all lower case). It allows you stop specific search engine from crawling some specific webpage(s) [7] [20] [22].

Syntax of basic robot.txt file is as follows [22]:

```
User-agent: *
Disallow: /
```

The declaration above stated "*" with User-agent, which is restricting all search engine not to index any of webpage.

User-agent: Googlebot Disallow: /

The declaration above will disallow/restricts Google crawler (Googlebot) from crawling and indexing entire website contents.

4.3 Off-page SEO Techniques

Off-page SEO techniques are those strategies that are nothing to do with contents of website. Unlike On-page SEO techniques, these strategies are done offsite and usually these are not visible on website [13] [21] [24]. Mainly, off-page SEO techniques are applied after implementation of On-page SEO techniques for improving website ranking [9] [23-26]. The process of applying off-page SEO techniques on website are not performed at once. Though, these techniques are applied gradually for better and long term result for ranking of website and for driving targeted traffic to your website [23]. Some elements of Off-page SEO techniques have important role for improving crawlability/Indexing of websites. In section below, I will discuss some of Off-page SEO techniques which should be considered for effective crawling and indexing of dynamic websites.

4.3.1 Directory Submission

The submission of websites to directories is one of a good way to get crawling on targeted website frequently as well as for getting fruitful traffic from directories [20] [22]. As I described in section 1.1.1 that directories allows submitting websites. These directories organize the submitted link into different categories and sub-categories. Although, there is a small number of people who prefer to find information through directories, but still they are good source to get traffic from directories when website is new and does not have much links to it [22] [23]. Moreover, search engine consider a website valid for crawling if it is found in directories rather than that if it found on irrelevant/poorly ranked website. Most of directories are free and allow deciding anchor text (section 4.3.2) of your choice to your website.

4.3.2 Anchor Text

Anchor text is a linked/hyperlinked text appears in the webpage which can be link to other website or another webpage of the same website. From anchor text both website's user and search engines get the relative information about destination webpage [7]. Generally, it is appear as underlined text and mostly in blue color.

Syntax example of anchor text is as follows:

For Free CD Download

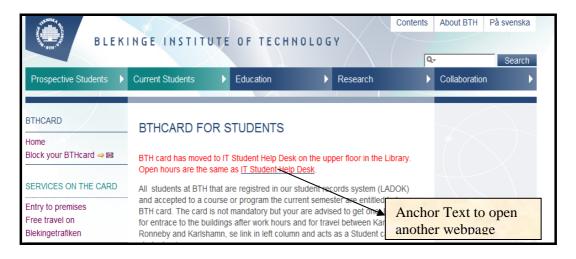


Figure 4.7 Example of Anchor Text in a webpage.

The anchor text reflects the targeted webpage/website contents. If anchor texts of website lead to related contents it results positive effect on the website impression to the search engine; conversely, if it leads to unrelated contents can be bad impact on website [13]. Therefore, use of targeted keyword or keyword phrase in anchor text is always a better option. But, there need to use that optimally by keeping keywords frequency and density in mind.

4.3.3 Link Building

Links building is very important element of off-page SEO techniques. During implementation of on-page SEO techniques, link structure mostly has out-bound links (which connect/link target website to other related and quality websites). These linking websites bring traffic to target website and bring positive effects on ranking of the website [7] [9] [13]. However, later during implementation of off-page SEO techniques outbound and inbound both types of links are built gradually. A successful link structure may take months [7] [22] [23]. There are not hard and fast rules that how many links are needed to get top ranking. The important thing that should be considered that target website is linking to quality websites (websites with high ranking). Reputed link (with better ranks) effect positively on another website and bad links can diminish website ranking [7] [9] [23] [26]. Therefore, this factor needs a good care to keep track of linking websites ranks, because a banned website (which is removed from search engine indices due to applying spamming tricks) can offer very bad effects on its linking websites (neighbor websites). The quality inbound and outbound links are good will be far more beneficial and successful. Another plus point for website linking to your website is that when ever spider aims crawling linking websites and it finds your website for crawling too [13] [20] [23]. Links building is categorized into two types of groups called as one-way link building (non-reciprocal linking) and two-way link building (Reciprocal linking). One-way link are those link which exit in target webpage but that link is not linking back to the target webpage; vice versa is called two-way linking [13]. Both types of link are important in Off-page SEO strategies.

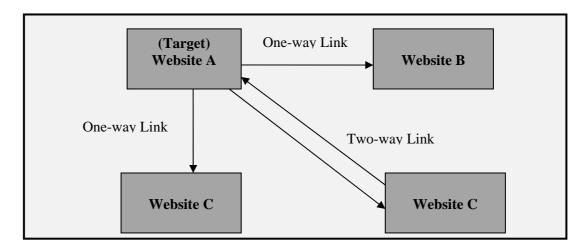


Figure 4.8 Example of one-way and two-way linking of websites.

4.3.4 Forums and Blogs

As search engine spiders like to crawl websites with unique and fresh contents. Therefore, frequently and regularly updated website contents ultimately results better indexing and ranking. Creation of blogs and forums allow website's users to enter comments and response against your website services and products [7] [20]. This feature allows website's users to come back to your website. This is beneficial in two different ways. First it increases website traffic. Secondly, it keeps webpage content fresh and updated which boosts crawling and ranking.

4.4 Important Design Considerations for Search Engine Friendly websites

The following section discusses some of the elements which become a need of website when it is needed to make easy navigation as well as easy interaction with website. On the other hand, these elements can cause obstacles in the attempt towards usability and their search engine friendliness of website [7] [13] [20] [22] [23]. For those reasons, they need attention to optimize them if it is obliging to use elements in website design.

4.4.1 Optimizing Frames

Frames in website facilitate website's users to see/find contents of several webpage(s) on a single (current) webpage; this feature reduces trouble of following links; moreover, a deep hierarchy to read target page might risk that website's users will come back to the current page. In contrast, use of frame in webpage gives tough time to search engines spiders; as, frames do not correspond to the web's conceptual model [7] [13] [20] [28]. Typical a webpage conceptually presents only one URL. Conversely, webpage that contains iframes or frames aims to display contents of more than one URL (one URL for each frame) inside a single webpage. This problem pose poor impact on webpage indexing; because, search engine spider cannot follow frames contents and it might miss important contents (webpage) within frame(s) [20] [28].

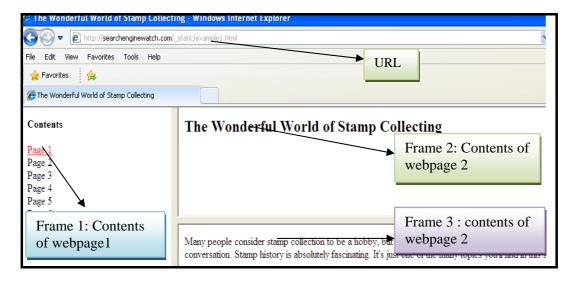


Figure 4.9 Example webpage that is using three frames to present 3 different webpage(s) contents to search engine against a single URL [28].

This example webpage aim to display contents from three different webpage(s); therefore, a single webpage URL is displaying content from three different webpage(s).

Google considered as rather smart for handling different types of web. It claims that Googlebot can handle "frames" and "iframes" but they do not guaranty that to what extent they have their grip to cope with frames [28] [30]. Therefore, search engines recommend avoiding frames in websites. Moreover, a website using frames often pose display problem for browser that does not support frames.

If there is no other way to avoid frames then search engine recommends using alternative of framed webpage(s). The alternative of framed webpage(s) is <noframe> tag that should contain exactly same contents of <frame>; otherwise, website might be consider to spammed by search engines [22] [28]. Search engines follow <noframe> tag and ignore actual frames. As <noframe> tag present a standalone webpage against each link/URL of frame; but it may cause navigation problem if that each webpage does not has link back to the home page. Moreover, it can cause search engines and user the no way back to the navigational page/home page. This might results that site will not be properly indexed by crawlers. However, a little effort can be made for optimizing framed website friendly for search engines; and accessible on browser that does not support frames [28].

Syntax of "frames" website and alternative "noframe" syntax of same website is provided below [28]:

Frame Code	Noframes Code
<html></html>	<html></html>
<head></head>	<head></head>
<title>Title of page goes</td><td><Title> Title of page goes</td></tr><tr><td>here</title>	here
	<frameset></frameset>
<frameset></frameset>	<frame src-<="" td=""/>
<frame src-<="" td=""/> <td>"navigations.html"></td>	"navigations.html">
"navigations.html">	<frame src-<="" td=""/>
<frame src-<="" td=""/> <td>"detail_page.html"></td>	"detail_page.html">
"detail_page.html">	<noframes></noframes>
	Alternative of frames Code
	/HTML code

4.4.2 Optimizing Forms

Currently, a large number of businesses rely on online buying, registration for products/services or ordering of product. Forms are popular mean of to provide interaction with websites. "Contact Us" is a simple example of webpage with form. Website's users need to fill out and submit forms. Later, on bases of filled form the resultant webpage is generated.

Unfortunately, conventional search engines' spiders cannot fill those forms [13] [20]. As a result, contents behind these forms remain inaccessible to the spiders. The contents of resultant webpage of formed webpage might be important to index.

Search engines recommend optimizing forms in two different ways i.e. by providing text link to the contents behind the login form (only for spiders). Other way is to implement "First Click Free" (FCF) [22] [29].

The FCF allows presentation of restricted contents in search index. It can be implemented in a way to allow searchers to see complete contents even they are registered or subscribed. However, when user clicks on original content it could ask for registration in an honorable way. Many famous websites being use this way to index hidden form's contents by search engines [7]. Google claim that it can fill forms but it still recommends that do not present registration forms to Googlebot instead make use of FCF for allowing crawlers to access restricted webpage(s) [29].

4.4.3 Optimizing Flash and JavaScript

The use of Flash and JavaScript is very popular among website developers. Flash is used to create animated graphics. Web-developers most likely use them to make their website attractive [13]. Sometimes, developers use Flash and JavaScript to create navigations which is not good for crawlers' point of view; because, Flash and JavaScript are troubling format for crawlers [7]. Flash navigation is not spiderable and results with the problem to crawl other webpage(s). Moreover, a website that heavily relays on Flash may cause website's user irritated; since, it take enough time to open webpage with excessive Flash items [7] [13] [22] [23].

Better option is to entirely avoid using Flash but sometimes websites' owner wants to have them in her website. Though, Flash can be coded in HTML [13]. It is possible to cope with this problem in favor of site owner, website's users and crawlers by making different layers of such webpage i.e. one layer visible for website's users, layer visible for search engine or browsers not capable to execute flash (mobile browsers). It is implemented by using <noscript>[13] [22]. However, <noscript> can be spammed by search engine if it not properly used. It is needed to take care to use same content for each layer (spider-visible layer and website's user-visible layer) [7].

Search engines completely avoid JavaScript to crawl; although, <noscript> is helpful because it gives alternative contents to search engines against JavaScript.

Syntax of noscript is as follows:

```
<html>
<head> head goes here</head>
<body>
<script type= "txt/JavaScript">
document.write("This is scripting world")
</script>
<noscript> Alternative of Script goes here">
</noscript>
</body>
</html>
```

Flash based webpage design need careful attention because too heavy webpage might not please website's user and search engines. In newer version of Flash, it is considered to make Flash contents discoverable to spiders [20] [22].

Another way to optimize webpage with Flash is to provide alternative HTML webpage against flash webpage. This solution will be rewarding in two ways. First, alternative webpage (HTML based webpage) will satisfy website's user; who does not like heavy Flashed webpage. On the other hand, search engine spiders would prefer crawling HTML version [22]. No matters that it need more work to do but it would rewards with better crawling. Recently, there are some tools (like swf2html) available to generate text against Flash content and links for Flash file [7].

4.5 Encoding URLs for Dynamic Websites

URLs are entering door to webpage for search engines and website's users. Therefore, SEO experts/web developers are convinced to optimize URLs by making them short, simple and readable for both website's users and search engine point of view. The optimization of URLs become more desirable and challenging in case of dynamic websites/webpage(s) because conventional search engines cannot interpret complex URLs with special characters [7] [13-16]. There are some techniques for URL rewriting and redirecting these important to consider for both static and dynamic websites under some circumstances [7]. The subsection below is provided with the brief details of some technique for URL encoding.

4.5.1 Redirecting

Redirecting becomes necessary when contents of one website/webpage are moved to other location i.e. website/webpage [7]. For instance, when contents of webpage http://www.domainName.com/old-page.html has been moved to new webpage http://www.domainName.com/new-page.html has been moved to new webpage http://www.domainName.com/new-page.html. Though, use of redirects indicates both web browsers and search engine spiders that content will be fetched from new URL. Whenever, old webpage (old-page.html) be followed, it will be redirected to new webpage (new-page.html). Without redirects both website's user and search engine crawler will get 404 error (Page not Found Error). The page not found error will result the loss of traffic of website. Moreover, this problem restricts search engines to find your webpage contents. In SEO campaign this problem can be quite paying in both cases. Therefore, redirecting is important to tell crawlers and website's users that where contents are moved for linked URL. Redirects can be used in websites in the following situations [7] [22]:

- Hosting company is changed.
- CMS is changed.
- Old contents that are expired can be move to new.
- Site has broken links but has traffic and other links.
- Needed to implement canonical redirects (robot.txt is discussed in section 4.2.12).
- Contents of target URL are moved to other URL for some reasons.

There are many ways to implement redirects. Generally they are divided into two main categories to direct crawlers, human and browsers to the new location (URL).

301 (permanently move)

This type of redirect informs both the browser and search engines that the contents of target resource have been permanently moved to the new location and there is no intention to bring them back to old location/URL.

302 (temporarily moves)

This redirect specifies that the contents of target resource have been temporarily moved to the new location and that is intended to move back on original location (URL).

However, crawlers interpret these redirects in a different way. When a search engine spider finds 301 redirect, it considers passing old link's traffic juice to the new link. Moreover, it will replace old link/URL with new link/URL in search engine indices. Conversely, when a search engine spider finds 302 redirect, it does not assume to pass link juice to the new link [7].

4.5.2 Methods for Redirecting

Redirects can be implemented in different ways. The subsection below is presented with some of standard redirecting method.

Redirecting methods vary according to web servers. Apache web server which runs on Linux or UNIX, redirects can be implemented by using a standard files known as .htaccess and RedirectMatch directives. On the other hand, web servers running on Microsoft IIS, allow to implement redirects by different method. A common method for redirects is employed for using IIS console (for Microsoft IIS) [7].

However, redirecting is possible to implement directly by using PHP, ASP, JavaScript and Perl. In each methods one thing is always need to keep in mind that redirect method returns in a way that web server returns 301 or 302 (HTTP status code) [7] [30].

The redirects that occur at page level can be implemented by Meta Tag.

Syntax of Meta tag for redirect might looks like this:

```
<meta http-equiv"refresh" content="4";
url=http://www.doamin.com/new.html"/>
```

Here "4" is specifies in first parameter of "content" that means the web server should wait for 4 seconds before redirecting to the new location. Search engine spider treats Meta tag redirects as "302 redirects". Though, parameter value content="0" will be treated as "301 redirect" by search engine.

In this research work, I aimed to work only with Apache server (PHP) language; therefore, I will discuss only related methods in brief detail [7].

.htaccess redirects

.htaccess file can be written in notepad. This file does not have name but 8 letter extension it may pose problem while saving file because some system like Windows 3.1 might not allow you save a file without name. However, this can be managed by saving file with any other name and upload on server later it can be rename by using FTP server.

Example of redirects code of .htaccess file:

```
Redirect /index.html /homes.html
Redirect /private/ http://www.newdomain.com/private/
Redirect /imges/picl.gif
http://www.domainname.com/images/picl.gif
```

RedirectMatch

This method for implementation serves and facilitates same as .htaccess does. But unlike .htaccess, the DirectMatch method makes use of regular expression instead of straightforward prefix matching [7]. The provided regular expression is matched against the

URL, on case of matching, the server use the new URL as file name. The example below shows that if it is needed to redirect all GIF to the same named JPG files on www.newdomain.com.

The syntax of such code might be written like this:

```
RedirectMatch (.*)\.gif$
http://www.new.com$1.jpg
```

4.5.3 URL Rewriting

Simple and understandable URLs convey information about target webpage contents. A descriptive URL of webpage could be fruitful in two other ways. At one end, descriptive URLs help the website's user to get a clue that what target webpage going to deliver her. On the other hand, it also leads to search engine for better crawling of website. Both of these facts could lead website with more traffic and better crawling.

URL rewriting is can be used in following circumstances [7] [22] [14] [30]:

- Contents of webpage are moved from one location to another location.
- CMS is changed
- Hosting company is changed
- It is needed to make clean URL for minimizing useless parameters (session IDs) to avoid duplicate content issue.
- Need to make search engine and user friendly URLs

4.5.4 Method for URL rewriting

There are different tools available for URL rewriting. For Apache web server, you would need to place "rewrite rules" in .htaccess file for URL rewriting. Conversely, for IIS server you would need to use ISAPI plug-in [7].

The subsection below is provided with a short description of .htaccess for URL rewriting.

.htaccess for URL rewriting

For URL rewriting .htaccess file will start from the following coding lines:

```
RewriteEngine on
RewriteBase /
```

After this statement, rewrite rules are implemented. Like DirectMatch, it also used regular expression to matching [7]. The example below aims to give product name against product id in URL. Rewrite rule may look like this:

```
RewriteRule ^prod/([A-Za-z0-9_-]+)\htm$ productpage.php?pro_id=$12)
```

4.6 Overview of Three Major Search Engines and their Indexing Considerations for Dynamic Websites

A search engine is used frequently when someone is looking for information in WWW. Some search engines i.e. crawler based search engines, index most or all sites. Google, Yahoo and Bing are considered as major and frequently used search engines among others [17]. These search engines brings significant amount of the traffic to websites [6] [9]. Therefore, SEO efforts are commonly applied to websites by keeping these major search engines in mind. However, these search engines differ from each the others according to their algorithm and ranking factors [22] [30] [32] [33]. Google is considered as most popular search among all other search engines; because, it provides comparatively better search result and covers a vast market share [17]. Most of searchers prefer to use Google on other search engines when searching information through on internet [9]. The different search results against same search query shows that all major search engines use different algorithm for ranking and indexing of websites.

These major search engines serve more than just searching tool. They provide some extra features like news, mapping, and email; along with that different types of applications are also offered from these search engine companies. These additional features attract growingly amount of users to use their search engine.

According to market research report, published by comScore, Inc. in November 2010, Google is at top with the explicit core search market with 66.2% market share, following by Yahoo! 16.4% and Microsoft's Bing with 11.8% Ask.com captured 3.6% of explicit core searches, followed by Aol.com with 2.0% [17].

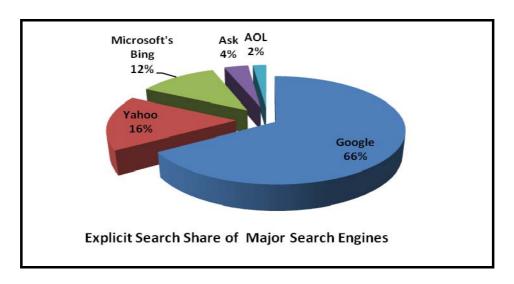


Figure 4.10 Explicit search share of major search engines.

These statistics demonstrate that Google, Yahoo and Bing cover around 94% of searcher preferences. My research work is aiming to explore SEO techniques for dynamic websites on Google, Yahoo and Microsoft's Bing. In others words, my research work will determine 94% of searchers which are affected by dynamic websites indexing problems in some way.

The following section aims to provide overview of these major search engines along with their indexing considerations for dynamic websites will be provided.

4.6.1 Overview of Google (Search Engine)



Google began in January 1996. It was result of research project by Lawrence Page and Sergey Brin; two PhD students at Stanford University, Stanford, California, USA [41]. The name "Google" was originated from the misspelling of word "googol" which means "A very large number. In other words, Google means "a numeral with one hundred zeros after it" which was destined to indicate the amount of information the search engine was to handle. In the beginning Google ran under the Stanford university website, with the domain google.stanford.edu. The Google appeared with registered domain name on September 15, 1997 and the company was incorporated officially on September 4, 1998.

According to comScore, Inc. currently Google is leading search engine with the 66.1% market share [17].

Google crawler is the software that is known as Googlbot [7] [30]. It is considered as extremely mature in searching most kind of web [7] [22]. Googlebot algorithm determines that how many webpage(s)/websites would be crawled during each crawling visit [30]. Google does not accept payments for increasing crawling frequency. The Googlebot explores available websites and stores the brief description of this website in search engine databases (indexes). When user enters search query in search area and hit the search button or press enter, it provide most related search results by extracting information from indexes.

The relevancy of contents is determined by more than 200 factors [7] [14]. Later, Rank is assigned these indexed webpage(s) on bases of these factors. The search results in SERP are

presented according to their rank higher. Highly placed result represent that it contains better ranking than the result in the bottom of SERP. Googlebot can process relatively more number of contents than other search engines, but not all types of contents like dynamic websites and media files may be indexed properly [13] [14]. However, Google provide some guidelines that what it consider while indexing dynamic websites [14] [30] [31].

4.6.2 Dynamic websites Indexing considerations of Google (search Engine)

In subsection below I will illustrate Google indexing considerations for dynamic website [22] [30] [34] [35].

Design and content guidelines

- **Site hierarchy:** Make website hierarchy clear and ensure that each webpage (desired to index) is reachable for crawler(s) at least once through a text based link.
- **Ready webpage(s):** Do not submit any webpage which is under construction.
- **Contents of website:** Provide unique and quality contents.
- **Links on website:** Links on webpage(s) (external linking websites) should be reasonable and relevant to contents of target website.
- **Title and Meta Description tag:** Contents under Title tag and Meta Tags should accurately describe the contents of target webpage.
- **Images:** Try use minimum number of images to describe contents of website. Use descriptive and reasonable names of images.
- Navigations: Avoid using images for navigation purpose because crawler cannot follow image for navigations.
- **ALT Tag:** Use ALT tag to describe what is image about.
- Non-friendly Contents: JavaScript, Flash, frames, DHTML, session _IDs, cookies
 might be trouble for Google search engines to crawl them. It is recommended
 avoiding or optimizing them.
- **File Size:** Keep file size reasonable. Google recommends using webpage (HTML based) of size 150kb.
- **Image format and size:** Keep the image size extremely small for fast download because images highly affect file/webpage size.
- **URL structure:** Keep URLs simple and smaller because it is easier to crawl.

Technical recommendations

- Make website discoverable: Submit website to Google, link it to other already
 indexed websites or submit website to directories. These factors make website
 discoverable for Google.
- **Submit Sitemap:** Google asks to submit sitemap (XML sitemap) to help Google to find webpage(s) (desired for indexing/crawling) for better crawling results. Google limits that sitemap should not exceed 5000 URLs.
- **Files and directories:** Keep files and images in organized way by keeping same types of files and images in same directories.
- **robot.txt:** Make use of robot.txt to direct crawler that which directories should/should not be crawled.

- CMS: Make sure that CMS created webpage(s), contents and links are crawl-able.
- **Test website appearance:** Check whether your webpage(s) appears properly in different types of browsers.
- Server response: It is needed to be sure that server response is good and it does not respond with 400, 401 and 500 errors. Time out/ slow server offer poor effects indexing.
- Page load time: Webpage load time should be as short as possible. It should not go over 9 second to load; because both crawling and traffic effected badly from page load issue.
- Well formed HTML Tags: Make it sure that the HTML code you are using is well formed i.e. all paired HTML tags are closed, and correct webpage with correct link/hyperlink.

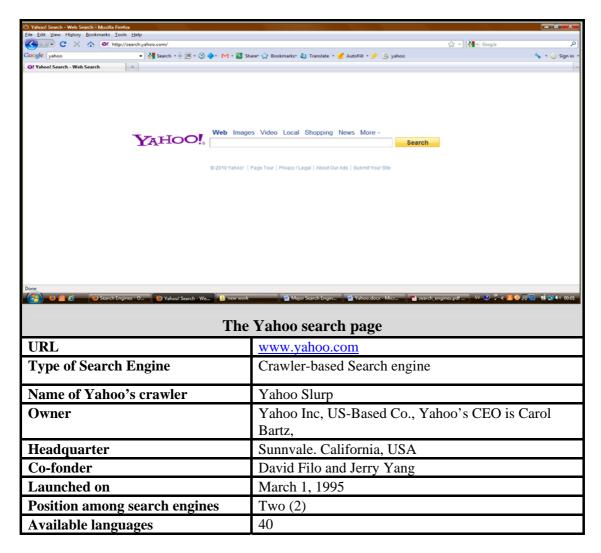
Quality measures

- **Avoid:** Clocking (do not present different contents to users and search engines), keyword stuffing (keyword density should be 2% to 3% of a page), bad neighbors (bad links) and doorway webpage(s).
- **Frequency of Crawl:** Frequency of crawl also depends upon factor like number of parameters in URL, rank of website and links to the page.
- **Links on website:** Link only those webpage(s) to website which have quality and relevant contents. Do not excessively link (external links) your website spider can consider it as spam.

Dynamic webpage(s)/URLs considerations

- **Dynamic URL parameters:** Google claims that it can deal with large number of parameters and deep webpage(s) but it recommends keeping parameter short and few in numbers.
- **Rewriting of URL:** Try to serves dynamic URLs because Googlebot can deal with them. However, it is better to remove unnecessary parameters instead of making it static looking URLs because it is hard to manage rewriting.
- Static link vs. Dynamic link: In some cases it is safer to present equivalent to static webpage(s) for dynamic webpage(s). However, other search engine are cannot interpret dynamic URLs; so try to make them keep URL parameters short and less in number.
- **Webmaster Tool**: Google webmaster tool allows specifying which parameters are not important in dynamic URL.

4.6.3 Overview of Yahoo! (Search Engine)



In 1994, Yahoo! was founded by two students, David Filo and Jerry Yang; they both were Ph.D. candidates in Electrical Engineering at Stanford University, California, USA [42]. They started their guide in campus trailer in February 1994 as mean to keep track of their interests on the internet. Previously, they were spending extra time on their home for managing lists of links. Ultimately, David and Jarry's list turned into enormously long and unmanageable, and then they decided to break the list into appropriate categories. When the categories turn into two, they decided to develop into subcategories. That was the inside theory behind birth of Yahoo's [42].

According to comScore, Inc [17], Yahoo is 2nd prominent search engine with the explicit core search market with 16.7% market share [17]. Yahoo is also a crawler–based search engine. Yahoo crawler is software and it is called as Yahoo Slurp [7] [32] that works almost same like Googlebot. It discovered and crawled new website if they are linked on other indexed websites. It also allows direct submission of website. Yahoo Slurp is not mature like Google because it face even more problems with dynamic website [7] [13] [15] [20]; therefore, it provide some recommendation about optimizing websites for better indexing of websites [32].

4.6.4 Dynamic websites Indexing considerations of Yahoo (Search Engine)

In subsection below I will demonstrate Yahoo indexing considerations for dynamic website [7] [13] [15] [32] [37].

Design and content guidelines

- **Site hierarchy:** Hierarchy of website should be clear; so that each desired page is accessible by Yahoo.
- **Ready webpage(s):** Present ready contents/ webpage(s) to end website's user and search engines.
- **Contents of website:** Contents of website should be unique and genuine with targeted keywords.
- **Links on website:** Link on website should intended to serve people for helping to find interesting and related information.
- **Title and Meta Description tag:** Contents used under Title and Meta tags should accurately describe the contents of page. Title and text in Meta Description tag should be between 25 and 150 and avoid using special characters in description tag.
- Images: Yahoo cannot understand what images are about; therefore, it recommends using descriptive names for Image files. Make minimum use of Image try use text to describe contents of page.
- **Navigation:** Make it sure that all desired webpage(s) have navigation with text link at least once to reach them.
- **ALT Tag:** Use ALT tag for images to tell search engines that what is image about.
- Non-friendly Contents: Frames, JavaScript, Flash, DHTML, session _IDs, cookies may be trouble for Yahoo like other crawlers to crawl them. Recommend to avoid or optimize them.
- **File size:** Keep file size reasonable like Yahoo recommends using page (HTML based) of size 150kb.
- **Image format and size:** keeping the file size small because they can cause page size heavier and slower to download.
- **URL structure:** Keep URLs of webpage(s) simple and smaller because it is easier to crawl.

Technical recommendations

- Make website discoverable: Submit website to Yahoo or link it to other already
 indexed website; so website will automatically be detected when those linking
 websites will be crawled.
- **Submit Sitemap**: Yahoo asks to submit site map to help it to finds webpage(s). Yahoo Sitemap limits up to 5000 URLs.
- **Files and directories:** Keep same types of files and images in same directories; so that crawler would find contents easily and efficiently.
- **robot.txt:** Make use of robot.txt to inform search engine that which webpage(s) should not crawl or index.
- **CMS:** If website is CMS based make it sure that created webpage(s), contents and URLs are search engine friendly.
- **Test website appearance:** Test your webpage(s) to check that they show correctly on different types of browsers.

- Server response: Make it sure that server does not response with 400, 401 and 500 errors because time-out and servers' slow response would badly affect crawling and indexing of website.
- Page load time: Page should be as shorter as possible it should not go over 9 second to get load. Because both crawling and traffic effected badly from page load issue.
- Well formed HTML Tags: Make it sure that the HTML code you are using is well formed and all tags closed at proper places.

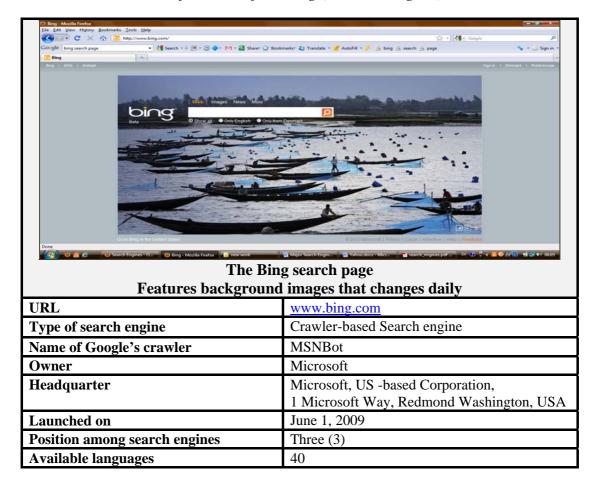
Quality measures

- **Avoid:** Clocking, keyword stuffing (keyword density should not be 3% in body and 15% to 20% for Meta description tags and in title tag) and bad neighbors (bad links) and doorway webpage(s) [13].
- **Frequency of Crawl:** Frequency of crawl may depend upon factors like number of webpage(s). Static website crawled frequently than dynamic website [15].
- **Links on website:** Website should not link to malware contents. Do not link website with irrelevant links/websites and do not links overly.

Dynamic webpage(s)/URLs considerations

- **Dynamic URL parameters:** Keep parameter short and few in numbers.
- **Rewriting of URL:** It is always good to remove unnecessary parameters from dynamic URLs; because it saves crawling time of spider to crawl unnecessary and duplicate webpage(s). It also reduces the chances of spider trap.
- Static link vs. Dynamic link: Yahoo claims can index dynamic webpage(s), but for page discovery it prefers static over dynamic page. It recommends avoiding dynamically generated links/URLs.
- **Webmaster Tool:** Yahoo site explorer also gives feature to rewrite URL to make them search engine friendly.

4.6.5 Overview of Microsoft's Bing (Search Engine)



Microsoft launched MSN search in 1998 as MSN Search. In the beginning, Microsoft did not get serious about improvement of search until after Google establish the industry. Initially, they relied on their partners like Overture, Looksmart, and Inktomi to control their search service. They started their own technology sample of their search engine in January, 2004 [20] [43]. They are officially powered from Yahoo!'s search results to their own internal technology on January, 2005. On May 4th, 2006 MSN declared their untenanted Yahoo!'s search ad program. After that Microsoft announced they were introducing their own live search product. Finally, On June 1st, 2009 Microsoft launched Bing [43].

Latest market research report [17] shows that Bing is at 3rd place among other major search engine with the explicit core search market with 11.2 % market share [17].

Researches reflect that Bing searches might be provided with fewer results against search queries than other search engines, but they usually have higher conversion rates. Therefore, making websites friendly for Bing can more rewarding that efforts it need.

After evaluation of Bing, it improved its user interface (UI) for search more quickly and easier ways. Until now, it improved its engineering work on a large scale. Now it crawls and index different types of contents. Later, it applies appropriate algorithms to finally present most search results (SERPs) to the searchers.

It works same like other crawler-based search engine by crawling, indexing websites and ranking them according to their contents. The Bing's crawler is known as MSNBot [13] [43]. MSNbot index and ranks better if website is having updated and quality contents. Bing has provided a guidelines and techniques to get better indexing [16] [33].

4.6.6 Dynamic websites Indexing considerations of Microsoft's Bing (Search Engine)

In subsection below I will discuss Bing indexing considerations for dynamic website [20] [16] [33] [38].

Design and content guidelines

- **Site hierarchy:** Keep your website hierarchy fairly smooth i.e. each target webpage should reachable within three clicks away from the main webpage.
- **Ready webpage(s):** Do not present under construction page to crawler and website's users.
- **Contents of website:** Contents of website should be unique and genuine and try to cover one topic for on one page.
- **Links on website:** Link on website should intended to serve people for helping to find interesting and related contents. Check carefully that website does not have link with malware software.
- Title and Meta Description tag: Use only few target keywords or keyword phrase
 contents used under Title and Meta tags should be accurately describe the contents of
 page. Title text in Meta Description tag should be between 25 and 150 and avoid
 using special characters in description tag. Make use of unique keywords for each
 title tags and Meta tags on each page.
- **Images:** Try to use text to describe contents, keywords instead of image i.e. do not use image embedded text that is needed to index. Minimize use of images and use proper names of image files because search engines cannot interpret images.
- Navigation: Make it sure that all desired webpage(s) have navigation with text link at least once to reach desired webpage(s). Avoid using images for navigations.
- **ALT Tag**: Use ALT tag for images to inform search engines that what is image about.
- Non-friendly Contents: Almost every search engine face hard time with frames, JavaScript, Flash, DHTML, session _IDs, cookies might be trouble for Bing like other crawlers to crawl them. Recommend to avoid or optimize them.
- **File size:** Keep file size reasonable like Bing recommends using page with no images should be of size 150k.
- URL structure: Keep URL simple and smaller because it is easier to crawling them.

Technical recommendations

- Make website discoverable: Submit website by using Bing webmaster tool or by linking other already indexed websites to make it discoverable for Bing.
- **Submit Sitemap**: Bing recommends submitting site map to help it to available all desirable webpage(s) through sitemap. Sitemap may contain up to 5000 URLs.
- **Files and directories:** Organize related files and images in same directories to make it easier to find contents and efficiently for crawling.
- **robot.txt:** Make use of robot.txt to prevent search engine from crawling or indexing specific files or directories.
- CMS: CMS based website need some attention that whether CMS created webpage(s), contents and URLs are search engine friendly.

- **Test website appearance:** The best way to be sure that your site is usable and appear properly in different browsers.
- **Server response**: Try to reduce server load by implementing page compression techniques and Conditional GET. Make it sure that server does not response with 400, 401 and 500 errors because time-out and servers' slow response badly affects crawling.
- Page load time: Page should be smaller in size; it should not exceed 9 second to get load. Because both crawling and traffic effected badly from page load issue. Both crawlers and website's user does not wait later than 10 seconds for webpage load.
- Well formed HTML code: Make it sure that HTML code is well formed and all paired tags are closed properly.

Quality measures

- **Avoid:** Avoid keyword stuffing, cloaking, hidden text, bad neighbors.
- **Frequency of Crawl:** Frequency of crawl may depend upon number of factors like website depth. Also website has more likely unique and quality content may crawl frequently than others. Static website may crawl frequently than dynamic website [16].
- **Links on website:** Links on webpage(s) should be relevant and should not overly use.

Dynamic webpage(s)/URLs considerations

- **Dynamic URL parameters:** Keep parameter short and few in numbers.
- Rewriting of URL: Make sure use simple URLs with targeted keywords. It is always
 good to remove unnecessary parameters from dynamic URLs; because it saves
 crawling time of spider to crawl unnecessary and duplicate webpage(s). It also
 reduces the chances of spider trap.
- Static link vs. Dynamic link: Bing suggests keeping URLs simple and static because they easier to crawl for search engine's spider and easy to remember too.
- Webmaster Tool: Unlike Google and Yahoo, Bing webmaster tool does not provide feature for parameter specification i.e. which should not be consider in dynamic URLs.

4.7 Useful Search Engine's Tools for SEO Campaign

There are several tools available for SEO campaign. Some tools are free and offered from search engines like Google, Yahoo, and Bing [34-39]. In this section, I will discuss some most popular and frequently used tools among SEO experts for SEO campaign.

4.7.1 Google's Useful Tools for SEO

The subsection below is provided with tools presented by Google for SEO.

Google Webmaster Tool

Google Webmaster tool is free for webmasters for SEO campaign. It helps webmasters to get a better control on website indexing and ranking [34]. This tool provides web developers with detail information about websites that how Google interacted with their websites. It also

presents handy information about websites like contents, Meta tag contents are either too long or short; as well as information about contents duplicated or repeated to many times [7] [20] [34]. Moreover, it allows finding out what search queries lead searchers to specific webpage. However, this tool does not guide how to improve websites to get better ranking. With Google Webmaster tool following services can be availed [30] [34]:

- Upload Sitemap.
- Generate and Analyze robot.txt.
- Which webpage(s) have problem to crawl through them.
- Broken Links.
- Find out search queries used to lead searcher to any specific webpage(s).
- Identify issues related to Title, and Meta description Tags.
- Get a clue that how Google crawler interacted with website.
- Remove desired webpage(s) from Google search result.
- Other notifications about violation of Google guideline like file size, malware.

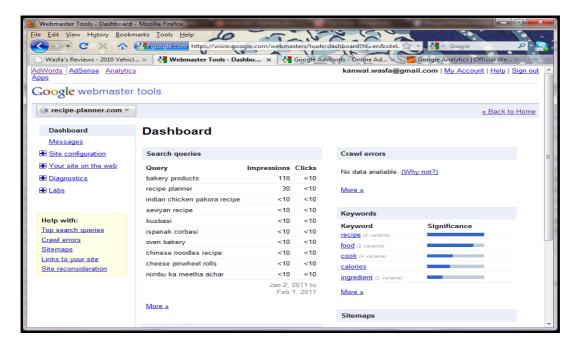


Figure 4.11 Google Webmaster Tool screenshot.

Google Analytics Tool

Google Analytics Tool provides facility to look information about incoming traffic of website. This tool offers following services [35]:

- Facilitate to judge impact of any optimization effort on website.
- Detailed information about how searcher arrive at website and what actions performed by searcher on website.
- Discovery of most well-liked contents or webpage(s) of target website.

Recently, Google enchased Google Analytics features with joint collaboration of Adobe named as "Google Analytics Tracking for Adobe Flash". This featured Google Analytics tracks that how website's users behaved and interacted with Flash contents of website [7] [22] [30] [31] [35]. This feature allows event tracking of Flash contents (online games and embedded videos and branded microsites), pageview [Google link new] of website.

However, Advance users are provided with wide range of information to improve website design and contents [31] [35].

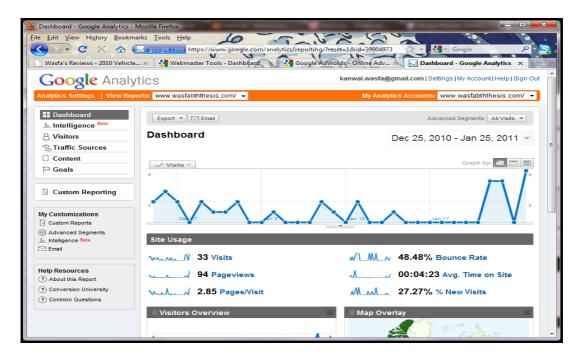


Figure 4.12 Google Analytics Tool screenshot.

Google AdWords

AdWords is Google tool [36] for flagship advertisement of products and services. Business that does not want to go for SEO to become visible in search engines (paid results) use Google AdWords tool for businesses promotions. Business pay for advertisement but in return they do not have to wait to appear in the search results (Paid results) Google's AdWords program allows businesses to pay to appear in search results [36]. It includes national, local and international advertisements. It offers PPC advertisement where advertisers choose the keywords that want to trigger for their ads. These result displayed separately from the organic results or at right side of first SERP in highlighted box (colored box mostly blue). Google order these paid results on bases of other advertisers' bidding and quality contents [3].

The Google AdWords campaign offers following benefits [3] [7] [13] [20] [36]:

- After the ads is submitted it appear immediately in search result unlike organic result does not need to wait for indexing in Google database).
- Changes in Ads are possible any time; it allows changing keywords to boost more to grow business.
- Google AdWords also have keyword tool. It allow keywords discovery for PPC campaign. This tool helpful especially in start of ad promotion. It also help when it needed to find better keywords when chosen keywords not much rewarding.

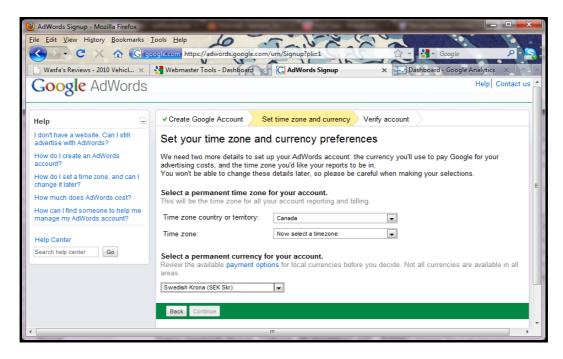


Figure 4.13 Google AdWordsTool screenshot.

4.7.2 Yahoo's Useful Tools for SEO

The section below is provided with tools for SEO for Yahoo Search engine.

Yahoo site explorer

Yahoo site explorer provided almost same feature like Google webmaster [37]. This tool allows the webmasters to explore all the webpage(s) indexed in Yahoo search engine. It makes it possible to following services [32] [39]:

- Submit website.
- Explore indexed webpage(s).
- Reasons that why any page is not indexed.
- Submit Sitemap.
- View most popular webpage(s) from any website.
- View comprehensive Sitemaps.
- Find linking webpage(s) to any popular website.
- Allows you to add "Site Explorer Badge" to website or any specific webpage. This
 feature allows you to track "inbound links" count to website and webpage
 respectively.
- Add Dynamic URLs
- Delete URLs from Yahoo indexes.
- Authentication to your website allows authorized access to the detailed information about website.

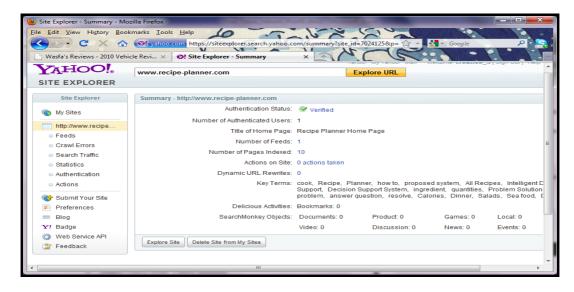


Figure 4.14 Yahoo site explorer screenshot.

YSlow for FireFox

Yahoo's YSlow is an Add-On [7] [13] [32] for Firefox that provides functionality to optimize website components including scripts, images and style-sheets etc. This useful tool is available for free. This add-on is incorporated with Firebug web development tool. It extracts the information related to components of website i.e. size, expired headers, and website is zipped, gzipped or not [32]. It grades website based on some user defines rule sets.

This tool uses tool to optimize images of website. It squeezes last few bytes of image to optimize performance without affecting the actual image. It also suggests how many bytes can be removed to save space and improve site speed. It suggests ways to improve site performance.

If provide following information related to website [13]:

- It summarizes webpage components.
- Display webpage statistics.
- It analyzes webpage(s) and explores that why they are slow.
- Provide tools (Smush.it and jSLint) for analyzing performance.



Figure 4.15 YSlow for FireFox screenshot.

4.7.3 Microsoft's Bing's Useful Tools for SEO

Like Google and Yahoo; Bing also provides tool for website's SEO known as Bing Webmaster Tools [20] [33] [38]. The subsection below briefly discusses description of Bing tools for SEO campaign.

Bing Webmaster Tool

Bing webmaster tool allows web developers to optimize their website for Bing search engine [38]. Bing updates its Webmaster tool to provide more precise information about indexing and crawling and traffic of website. This tool is also free of cost and allows availing following facilities for websites [20] [33] [38]:

- Submit Website
- Submit Sitemap.
- Submit individual URLs.
- Get detailed information to know which webpage(s) of website are index.
- Information about keyword performance, back-links and inbound links.
- Provide information like issues of crawling, malware, redirects and other exclusion encountered during website crawling.
- Provide charts to get a quick view of last six months traffic, crawling and indexing of website.



Figure 4.16 Bing Webmaster Tool screenshot.

4.7.4 WebRank Toolbar for Firefox

This tool is Mozilla Firefox add-on to provide Google Page-rank, compete, Alexa and Quantcast of website currently being views in browser [39]. Along with that it shows the information of indexed webpage(s) in major search engines (Google, Yahoo and Bing) [20] [32].

Below is a screenshot of WebRank Toolbar. It also allows getting detail view by click on button (shown with arrow) for search engine. For instance, detail view of Google will lead

you to view list of indexed webpage(s). It also provides back-link tracking facility for websites [39].

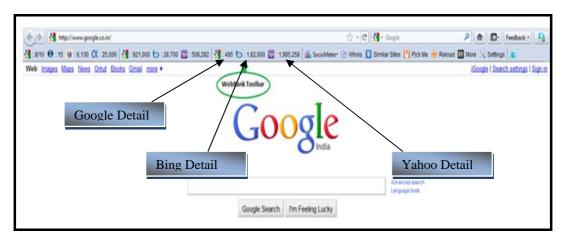


Figure 4.17 Webrank Toolbar screenshot.

5 EXPERIMENT

This chapter consists of goal definition, experiment planning phase, experiment operation process and experiment sample. In this chapter, I will discuss some experiment's theories with experiment planning and execution process step by step i.e. how the experiments are conducted in reality.

5.1 Goal Definition

The goal definition is the statement to define the major aspects of experimental study before proceeding planning and execution. The objective and purpose of study, quality focus, perspective, better understanding and filling research gap in exciting knowledge with meaningful impact upon the field of study are some major aspects that should be included in goal definition.

This experimental research is aimed to predict and determine effects of one variable on experiment by controlling several factors that may cause results to vary. My primary purpose is to use experimental research method to explore commonly used SEO techniques for improving indexing of dynamic webpage(s) on major search engines. Secondary purpose is to compare SEO techniques and see effectiveness of one technique over other.

Literature review shows that the dynamic websites are commonly known for poor visibility in search engines because they are not search engine friendly. There are several myths about dynamic website's visibility in search engines; and search engines are continually evolving to make them efficient, and some improvements are made in this regard. In this empirical study, I aimed to investigate how much advancement have been made in the area of indexing dynamic websites; and to identify if there is any problem in indexing dynamic webpage(s) in major search engines. Further, using SEO techniques to help in indexing these webpage(s) is optimum purpose of my experimental research. This research work leads to a better understanding; fill up a research gap in exciting knowledge and going to have meaningful impact upon the field of web development and SEO.

5.2 Experiment Planning

Experiment Planning is not an easy task; as a lot of preparation is required to plan and design experiments [43]. A well planned experiment ensures that an experiment is carried out appropriately and that the result reflects the real world, in the best credible method. This experiment planning phase has been designed based on guidelines provided by Wohlin et al. [43] Campbell and Stanley [44] Campbell and Cook [45].

5.2.1 Hypothesis

A hypothesis is an illumination for a phenomenon which can be tested in some way which idyllically whichever proves or disproves the hypothesis. Until testing of hypothesis, the hypothesis in use to be considered true and the objective of the researcher is to methodically test the term of the hypothesis. Commonly researcher use null and alternative hypothesis in the phase of formulation of hypothesis [43]. We can describe a null hypothesis as a statistical hypothesis; null hypothesis is affirmed for acceptance [43]. It is also called as original hypothesis. Any other hypothesis other then null hypothesis is known as alternative hypothesis [43]. When null hypothesis is rejected we must accept the alternative hypothesis. Null hypothesis is symbolized by H_0 and alternative hypothesis is symbolized by H_1 [43].

For this experiment, formulated null and alternative hypotheses are explained in Chapter 3 (section 3.1).

5.2.2 Selection of Variables

In experimental research, variable can be described as a measurement or an attribute, whose value can vary over the course of an experiment. It helps in identifying the cause (independent variable) and effect/outcome (dependent variable) in the term of an experiment [43] [45]. Wrong selection of variables can effect badly on results of experiment [43]. Therefore, it is very important to select correct variables [43]. There are two well-known variables i.e. Independent and dependent. An independent variable is the one being cause, manipulate or affect outcomes of an empirical study [43]. Conversely, the dependent variables are depending on the treatment behavior of independent variables. Dependent variables are observed and deliberated to fix on the effects of independent variables. In other words, dependent variables are outcomes of research study [43] [45].

In this experiment, I have selected following independent and dependent variables:

• Independent Variables:

SEO Techniques: Independent SEO technique or set of SEO techniques are independent variables that are used to help dynamic webpage(s) to get indexed on targeted search engines.

• Dependent Variables:

Dynamic Webpage(s) Indexing: Dependent variables would be the indexing of dynamic webpage(s) in targeted search engines.

5.2.3 Selection of Subjects

In an experimental research, it is important to select right subject because selection of subject is directly associated to generalization of the results derived from the experiment [43]. In order to simplify the results to the selected population, the selected subject has to be representative for that population. Subject selection is also known as sample from population [43]. There are two types of techniques for population sampling: probability and non-probability. In probability sampling, the probability of selecting each subject is known whereas in case of non-probability selecting a subject is unknown [43].

In this study, I used probability sampling technique for subject selection purpose. I opted to experiment some SEO techniques (particularly friendly URLs) that are involved in indexing dynamic webpage(s) on targeted search engines. For this purpose, I decided to use two dynamic websites that of different in age and size and are uploaded on different servers. Main reason to use two sites was to see if search engines worth size, and age of a websites when indexing dynamic URLs. It was likely to doubt that results from one website are generalized to particular websites under study.

Sample of webpage(s) is randomly selected from webpage(s) that were not indexed on targeted search engines. It was not possible to experiment all SEO techniques and fully optimize a dynamic website due to limited scope of research; so I restricted my study to techniques which help indexing dynamic URLs. To ensure I am not working on indexing a website that's lacking in basic SEO techniques, I created a new website using On-page SEO techniques. Another reason to choose indexing dynamic webpage(s) as my sample subject is to study the most controversial areas of dynamic website (dynamic URLs), particularly dynamic URLs using parameters and session variables in their URLs which are considered a problem for most of search engines.

5.2.4 Experiment Design

Experiment design is a plan to accumulate experiential knowledge, i.e. knowledge based on the analysis of investigational data [44] [45]. It can be practical when ensure an occurrence sequentially to improve understanding or get better performance [44]. Making up a design means attentively choosing some experiments that are to be executing under controlled arrangements [43]. After understanding of guidelines and unified steps, which are involved for choosing an appropriate experiment design [43] [44] [45], I chose to use The Pretest-Posttest Control Design Group as my experiment research design. This design consist two groups i.e. experimental and control. Experimental group is given the treatment and the results are accumulated at the end of experiment. The control group obtains no treatment, over the same period of experiment time, but undergoes properly the same tests.

The experiment design representation:

If O₁ represents pretest of experimental group, X represents the treatment and O₂ represents the posttest of experimental group, then difference between O₁ and O₂ will be explained by X. Similarly, O₃ represents pretest of control group that would not be exposed to any treatment; however, this group will be tested again after treatment. X is only applied on experimental group.

Experimental Group: Pretest → Treatment → Posttest

 $O_1 \times O_2$

Control Group: Pretest → Posttest

 $O_3 O_4$

In experimental group of this experiment, O1 represents static and dynamic webpage(s), which are not indexed on targeted search engine. X represents SEO techniques, which will be applied to help dynamic and static webpage(s) to get indexed on targeted search engines. O2 represents results of applied SEO techniques. In control group, O3 represents static and dynamic webpage(s), which are not indexed on targeted search engine. O4 represents results without applying any treatment (SEO technique).

5.2.5 Validity Evaluation

In any study, validity of results is fundamental question for researcher. All research designs are likely to have bias and design threats which can affect validity of results [43]. Validity is the degree to which an experiment procedures what it claims to test. It is essential for experiment to be valid sequentially for the results to be precisely applied and comprehend [43]. In the beginning researcher classify two types of validity threats (Internal and External) [45]. Present research expanded into in four categories (Statistical Conclusion Validity, External Validity, Internal Validity and Construct Validity) [44].

In subsections below, I describe those validity threats, which can affect on selected research design of this experiment and how I controlled and addressed them.

External Validity

Experiment results and/or conclusion which can be validly generalized from a sample back to its parent population is said to have external validity [44]. Due to this validity threat experiment results may not able be generalized to real world. External validity is defined in terms of Population Validity and Ecological Validity. In this experiment, it is not difficult to generalize results of this study because results are not affected if they are performed at variety of locations and there are any participants that can affect experimental data.

- Population Validity: It is defined as the degree to which experiment finding and/or
 conclusions can be generalized from a sample back to its parent population [44]. It is
 possible to generalize from study participants (subjects) or the specific finding
 (conclusion) based on differential characteristics. However, this validity threat is
 under control because no human participants are involved in the study.
- **Ecological Validity:** It is concerns results generalization to other environmental conditions [44]. This factor is also under control because results of my studies are not affected by variety of placed where experiments are performed.

Internal Validity

Experiment results may not be accurate because something else besides treatment X affected results of the experiments. Internal validity threats can be factors that can affect the independent variable without researchers' knowledge and affect the cause [43] [44]. Some factors affect internal validity includes:

- **History**: Events which take place between observation and measurement, in addition to independent variable which affects dependent variable, are called history [45]. It usually happens when some unplanned events occur during the experiment. During in this experimentation, I kept a complete track of optimization. I used search engine webmaster tools and server files to keep track of any unplanned events. All factors were known and such incident when server was down for one of the websites and as a result of this problem crawlers failed to indexing website.
- Maturation: When subject changes between an interval T1 and T2. This often
 happens when people, especially children, performing a treatment on an on object
 change. It includes emotional, psychological, or physiological processes access time
 within study subject, usually individuals which operate on variables in somehow [45].
 Since it is not a study where human beings are not subject; so there was not any
 maturation threat.
- **Testing**: When data gathering exercise is repeated it is likely to affect subject performance or recall [45]. When same pretest and posttest are performed, results may affect. In my experiment, there was not repetitive pretest that could affect posttest; so this threat has no implication of validity of results.
- **Instrumentation**: Changes within test or data collection devise or raters may produce changes in scores on or measurements of the dependent variables [45]. If different measures and administrator are used for pretest and posttest, results are likely to be affected. In my case, administrator did not changed throughout the process and measurement method and standard remained same throughout the process. So this threat remained under control.
- **Statistical Regression**: If group of subjects and assembled based on extreme sores or other measurements results are likely to move towards group mean [45]. When subject started recording at high (outliner) on a spectrum and can only go down and vice versa. There were not any outliners in this experiment so results are significant.

- **Selection Biases** If experimental and control groups are not identical, it is likely that false variables may be affect results [45]. In this experiment, this factor was carefully considered and experimental group was selected randomly and controlled group webpage(s) and identical in terms on data organization and basic SEO techniques.
- **Experimental Morality:** The differential loss of study from research group is called morality [45]. Results are affected when some subject leave in the middle of a study. It is more likely to occur in long studies. In my case, since it is a short study so there was not any human subject so this threat remained under control.
- Resentful Demoralization of the control group: When control group elects to perform differently than normal, because it is not getting treatment like experimental group [45]. In this experiment, Control group and experiment group remain alike in terms of changes and there were not any resentful changes going on control group. Only experimental group was subjected to treatments.
- **Diffusion of Treatment:** It is likely to occur in blind studies due to differences within group; it may occur due to unintentional independent variable [44]. Since I did not perform a blind study, this threat was not significant to make any difference in results.

Statistical Conclusion Validity

It is concerned with the issue that affects the ability to draw conclusion about relationship between treatment (X) and outcome (O₂) [44]. Statistical inference evaluate whether data validation procedures produce correct results so there is a need to make assure author understand statistical results performed study.

In this experiment, I use SEO techniques step by step, one technique at a time by targeting particular search engines. The effect of techniques is already known when they are applied on targeted group and results are monitored on all search engines individually. This threat is mitigated because I have prior experience of doing statistical tests so it is not very likely that I will perform statistical tests incorrectly.

Construct Validity

It concerns with generalizing the results of the experiment to the theory or concepts [44]. According to Wohlin et al. [43], there are two kinds of threats to construct validity: threats related to design of experiments and threats that are concerned with social environment i.e. behavior of the participants. In this experiment, results are recorded from search engines after certain SEO techniques are applied; therefore, these results can be used to generalize the efficiency of SEO techniques for indexing of dynamic websites. As I mentioned in section 5.2.4 that Experiment design is selected after understanding of guidelines and integrated steps, which are involved for choosing an appropriate experiment design and I consider those validity threats, which can affect on research design of this experiment. Also, this experiment did not involved individuals to participate in study; so in these circumstances construct validity threat does not subsist.

5.3 Experiment Operation

The process of experimental operation can be described as the procedure of meanly defining variables into measurable elements [44] [45]. The procedure identifies fuzzy concepts and authorizes them to be measured, empirically and quantitatively. In this experiment, experiment operation phase includes Instrumentation, execution and data validation.

5.3.1 Instrumentation

In experiments, instrumentation is known as most time consuming practice. It comprises of selection of tools, documentation, and creation of guideline to perform experiments [43]. It also contains developing pretest and posttest questions. It is very important to carefully design pretest and posttest questions; because later in the study they are subjected to answer research questions.

For instrumentation of my research experiment, it involved quite some work setting up experiment. Since I selected two websites for performing experiments, one of the website had to be a new dynamic website so that I can closely monitor, how search engines respond to a new website in terms of indexing its webpage(s). I had to create a sample dynamic website for this purpose.

I also planned to examine effectiveness of URL rewriting on search engines. For this purpose I planned to use a websites where I can easily tryout URL rewriting. It is effective to use URL rewriting in early stages of website development because it is quite some work to implement URL rewriting on a developed website. Another reason to keep URL rewriting on a different website was to control any possible influence on newly uploaded website. So I used a website that was developed in PHP template engines which allows creating friendly URL with less effort.

Wohlin et al. [43] classifies three types of instruments, which are known as common for using in experimentation i.e. Objects, Guidelines and Measurement tools. Following of detail description of instruments which I used in this experiment:

Objects: In computer science experimental research, object can defines as specifications or code documents which are used to execute experiments [43]. Objects used in this experiment are webpage(s) that are subjected to indexing on targeted search engines. Also search engine's SEO techniques that are specific to dynamic websites and common for static and dynamic websites.

Guidelines: Before execution of experiment it is necessary for researchers that they are well aware about guidelines to execute the experiment. Also, they must need to know about topic and its background, aims and scope of experiment in depth [43]. For this experiment, I reviewed several research papers to familiarize myself with experimental research and writing technical documentation. I also performed research on how to perform experiments for investigative studies. After understanding of experimental research, I decided on scope of chosen topic and prepared guideline for myself.

Measurement Tools: The tool used for measurements and data collection on a variable is called measurement tool [45]. In this experiment, I used webmaster tools for Google, Yahoo, and Bing to perform SEO techniques. I also used Google Analytics account to keep record on traffic and monitor traffic reports. Other than these online tools, I also recorded search results from SERPs of each Search on daily basis and used excel sheets to record collected data.

5.3.2 Execution

After an instrumentation process has been done, the next step is execution of experiment. This experiment is conducted in a period of two month after development and optimization

(SEO) of both websites which, I am using in this experimental research. Initially, it was not known how long search engines will take to index webpage(s) (experimental group) because it depends upon search engine's results. However, I intended to complete experiment in a period of 30 days. In addition to experiment execution, I wanted to test effect of time factor on search engine indexing behavior. So, final readings were taken after a period of approximately one month after applying last SEO technique.

Since experiment required online websites so there was no need to have fix setting or place to take experiment readings. However, computer labs of Blekinge institute of Technology (BTH) and my home network were two places that I worked most of the time.

To avoid SEO techniques results overlapping, I applied and evaluated SEO techniques step by step. To record results of experiment I used online webmaster tools for Google, Yahoo and Bing. I also relied on search results generated through manual search queries on each targeted search engines. I used two websites in this experiment; one websites was new to the search engines and seconds has web presence since last one year. Main purpose to use two websites of different ages was to study, if search engines response to them differently or not.

First, I created a new dynamic website and uploaded it on server so that I can study behavior of search engines and study if search engines can index the target website without setbacks. Second, I applied SEO techniques, one technique at a time, on webpage(s); I was checking search engines' results every 6 hours. However, I was recording experiment results only once in a day. There was not any need to records results frequently because results were not very likely to vary but I was taking my reading according to my plan. There was only one instance that results were seen changed within six hours after performing one technique in Bing webmaster tool.

Results were recorded in Microsoft Excel sheets and manual search results were recorded as screenshots taken from webmaster tools and search engines.

5.3.3 Data Validation

The process of data validation is important to determine the technical usability of the analytical data. It ensures that every data value is correct and accurate in the experimental research. It was important to carefully record experimental data to avoid wrong conclusions. In this experiment, to perform SEO techniques on dynamic webpage(s), I ensured that webpage(s) have similar page layout to avoid possible interference in resulting due to different page layouts and content organization. Since I used dynamic webpage(s), this factor was not hard to control. Experiments data was recorded around same time and appropriate screen shots were also taken to keep record of results. So, recorded data collection process is carefully designed and carried out.

5.4 Experiment Sample

For experiment sample, I selected to present my work done for most controversial SEO technique called search engine friendly URLs. As I discussed in section 5.2.4, I selected one website to experiment search engine behavior for URL rewriting. Although this website was available online since last one year, it was indexed only in Google; however, website home page was searchable in Yahoo, Bing, and AOL. Website was not indexed in Ask until I manually submitted website to Ask search engine.

Pretest and posttest Screenshots are taken before and after applying URL rewriting and sitemap creation. Pretest and posttest screenshots are taken from Google, Yahoo, Bing, Ask and AOL search engines are presented in Appendix-A. The presented search results are taken by manual search process in targeted search engines and Webmaster tools where fitting to display. The search results are very possible to change if search is repeated because search engines are always in the state of updating index databases. These provided search results are for www.recipe-planner.com.

6 ANALYSIS AND INTERPRETATION

In this chapter, I will present statistical and empirical analysis of data collected in the experiment. In section 6.1, I will discuss measurement preface. Section 6.2, is presented with brief explanation of each SEO technique that I used during experimentation. In section 6.3, I present descriptive results; Further, I will perform result's analysis statistically and descriptively in section 6.4. Finally section 6.5 contains testing of hypotheses that I formulated in earlier 5.2.1.

6.1 Measurement Preface

Before I present experimental data I would like to discuss implantation model. The purpose of experiments was to apply some SEO techniques to help dynamic webpage(s) get indexed on listed search engines. There were some limitations that affected our test and their results and it is important to discuss them before I present actual data.

External Factor: As mentioned in chapter 5 (section 5.2.5) that there are some external factor such as external links and server side problem i.e. server down can affect crawl ability of search engines. It was not possible to eliminate external factors, but I kept a close eye on server log files and results to see variation in results.

Instability of measurements: Since search engines are constantly in motion and evolving, and they update index databases frequently, so it is obvious if any measurements are taken or repeated again, same results might not yield.

Measuring tools: Search engine webmaster tools are not capable to keep complete record of indexed webpage(s) so I had to rely on manual searches along with available features in online webmaster tools.

6.2 Applied SEO Techniques

I discussed SEO techniques in details in Chapter 4 (Section 4.1, 4.2, 4.3); here, I will briefly explain techniques that I used during experimentation.

Sitemap Creation: Using this technique, I created sitemap of the website and uploading on server.

Sitemap Submission: If a website ownership is authenticated for Google, Bing, and Yahoo. It is possible to manually upload a sitemap in webmaster tools. It is possible to manually submit sitemap to Yahoo, Google, Bing, and Ask.

Website Submission to Search Engines: I submitted website to over 20 search engines.

External Link Building: Website1 was indexed only on Google, for indexing it on other search engines I linked website on two other websites that had web existence since last many years.

Manual URL feeds to Yahoo: Yahoo Site Explorer provides facility to manually submit URLs to its webmaster tool. In addition to that it allows encoding dynamic URLs to some extent. So this technique involves submission of dynamic URLs to yahoo Site Explorer Tool. **Manual URL Submission to Bing:** This technique involves manual submission of URLs to Bing.

Internal Link Building on home page: It involved creating hyper links of the pages needed to be indexed; Links are created on home page. This technique is particularly significant for important pages of a website. Footer links and bread crumbs are now commonly used to promote webpage(s).

Friendly URLs: URLs are encoded to include search engine friendly text as part of the URL itself to make URLs more search engine friendly.

Time Element - 30 days: After applying last SEO technique I left website for a period of one month to see how results vary with time. During this period, I did not apply any further techniques but I was keeping an eye on server log to see if there are any server side errors or crawler related issues.

6.3 Results

It took about two month to perform experiment which started on 4th Dec 2010 after website development and results were last recorded on 22nd Jan 2011. Results were recorded in two steps. In first steps, I applied SEO techniques step by step and experiment results were recorded on daily basis. For this step results were last taken on 24th Dec 2010 after applying last SEO technique. Then there was wait period of about thirty days to study effect of time element on search engine's indexing mechanism. I did not apply any additional technique during this period; however, I was keeping track of server log files to make sure there are no underlying server side problems that may affect experimental data. In second step, final results were taken on 22nd Jan 2011 for targeted search engines.

Although results were recorded on daily basis, results presented here are summarized to demonstrate effectiveness of each SEO technique useful on targeted search engines. In the following sub-sections, I will present results for both websites.

6.3.1 Website1: wasfabththesis.com

Website1, a relatively small website, contains total 21 URLs (including 18 dynamic and 3 static webpage(s)). I aimed to index these webpage(s) on search engines by applying certain SEO techniques sequentially; website's pages other then targeted 21 webpage(s) were access denied by crawlers. At the end of experiment total 67% webpage(s) were indexed in Google, 48% webpage(s) were indexed on Bing and Yahoo, and 33% webpage(s) were indexed on Ask and AOL Search engine including all static webpage(s).

This website was uploaded on server on the day when I started experiment so that I can watch indexing process. Before applying a SEO technique, a sample of webpage(s) was randomly chosen from the webpage(s) that were not indexed. SEO treatment is applied on experimental group. The control group obtains no treatment, over the same period of experiment time, but undergoes properly the same tests. Some SEO techniques presented in Table 6.1 do not have control group. The reason behind not having control group for those SEO techniques is since these techniques need to apply specifically on all webpage(s) of website. The comparison of those SEO techniques which have both experimental and control groups is presented in Table 6.2.

Index summary of website1 is presented in Table 6.1 and measurements are visualized in Figure 6.1. Number presented in the Table 6.1 and Figure 6.1 indicates total recorded number of indexed webpage(s) on website1 after applying certain SEO techniques sequentially. For simplicity, indexing results are summarized in Table 6.1. Each entry represents number of webpage(s) indexed after applying step by step SEO technique on experimental group. Efficiency of each SEO technique is also calculated individually to determine how effective each technique is; as shown in Table 6.5 (Section 6.4).

Table 6.1 Webpage(s) indexing summary for website1 (wasfabththesis.com)

wasfabththesis.com index summary									
Applied SEO Technique	Google	Yahoo	Bing	AOL	Ask				
Startup	1	0	0	0	0				
Sitemap creation	2	0	0	1	0				

Website Submission to SE	1	0	0	0	6
External Link Building	0	1	1	2	0
Sitemap submission to SE	5	0	0	3	0
Manual URL feeds to Yahoo	0	0	0	0	0
Manual URL Submission to Bing	0	6	6	0	0
Internal Link Building on home page	3	0	0	0	0
Time Element - 30 days	2	3	3	1	1
Total webpage(s) indexed (out of 21)	14	10	10	7	7

The website1 is relatively small in size created with on-page SEO techniques.

Home page of website got indexed on Google on the day it was uploaded on server. However, both Yahoo and Bing were not able to index website until it was linked on an external website. All targeted search engines responded differently towards targeted website. Yahoo and Bing search engines rely on each other to index webpage(s). Google and AOL responded to sitemap submission more than Bing and Yahoo. It is found that results vary on all search engines with time. During experiment, hosting server was down for about 24 hours, which affected results on Google webmaster tool. As a result of this problem, crawler visits resulted in errors and I had to resubmit sitemap to all search engines. Figure 6.1 is descriptive representation of indexing throughout experimentation with step by step application of SEO techniques for website1 (wasfabththesis.com).

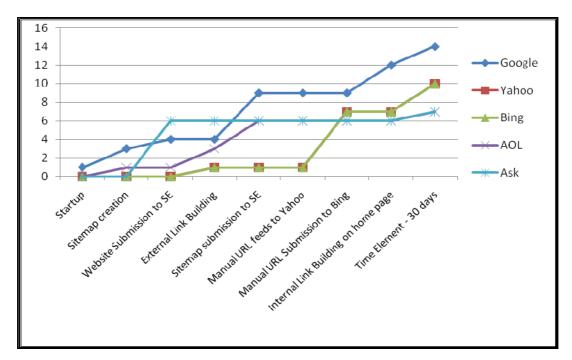


Figure 6.1 Descriptive representation of indexing throughout experimentation with step by step application of SEO techniques for website1 (wasfabththesis.com).

Table 6.2 contains comparison of experimental and control group for SEO techniques for website1 (wasfabththesis.com). Yahoo Site Explorer is very inefficient tool, URL submission to Yahoo is hardly a success results appeared to be Null for experimental group. Yahoo Site Explorer provided URL rewriting which is limited to URLs containing session key in it. Conversely, Bing Webmaster tool is very efficient to URL submission and results are seen in few hours after submission. In addition to that, URLs submitted to Bing are automatically indexed in Yahoo search engine. URL submission of dynamic webpage(s) is an efficient SEO technique for small to medium size websites. Internal link building and bread-crumbs

trail SEO techniques helped to index webpage(s) in Google. Webpage(s) that is not linked to any page remain orphan and did not be index on any of the search engines. So internal link building is an efficient local SEO technique. Comparison of experimental and control group for SEO techniques for website1 (wasfabththesis.com) is presented in Table 6.2.

Table 6.2 Comparison of experimental and control group for SEO techniques for website1 (wasfabththesis.com).

wasfabththesis.com index summary										
Applied SEO Technique	Google		Yal	hoo	Bi	ng	AOL		Ask	
	E	C	E	C	E	C	E	C	E	C
Manual URL feeds to Yahoo (6 URLs)	0	0	0	0	0	0	0	0	0	0
Manual URL Submission to Bing (6 URLs)	0	0	6	0	6	0	0	0	0	0
Internal Link Building on home page (3 Webpage(s))	3	0	0	0	0	0	0	0	0	0
Total webpage(s) indexed (out of 15)	3	0	6	0	6	0	0	0	0	0

6.3.2 Website2: recipe-planner.com

This website contains 131 dynamic webpage(s) with 38 webpage(s) containing friendly URLs which served as our experimental group to test effect of friendly URLs on search engines. This website site had web existence last one year with 18% webpage(s) indexed in Google. At the end of this study percentage increased to 57% on Google. Furthermore, a total of 20% webpage(s) were index in Yahoo, 18% in Bing, 15% in AOL and 13% were recorded in Ask. Results will be discussed in chapter 7 to provide more details about search engines' indexing behavior.

In website 2, some SEO techniques do not have control group. The techniques, without control group, are preliminary techniques needed to apply on all targeted webpage(s) of website. As I mentioned before, this website aim to test effectiveness of friendly URLs towards major search engines; therefore, this SEO technique have both experimental and control groups. The comparison of both groups is presented in Table 6.4.

Index summary of website2 is presented in Table 6.3 and measurements are visualized in Figure 6.2. The figures (number) presented in the Table 6.3 and Figure 6.2 indicates total recorded number of indexed webpage(s) on website2 after applying certain SEO techniques step by step. For simplicity, indexing results are summarized after applying step by step technique on experimental group. Efficiency of each SEO technique is also calculated individually to determine how effective each technique is; as shown in Table 6.6 (Section 6.4).

Table 6.3 Webpage(s) indexing summary for website2 (recipe-planner.com)

recipe-planner.com - index summary										
Applied SEO Technique	Google	Yahoo	Bing	AOL	Ask					
Start up	23	1	1	1	0					
Sitemap creation with URL rewriting	20	8	8	4	0					
Manual URL feeds to Yahoo	0	1	0	0	0					
Website submission to ASK	0	1	0	0	2					
Sitemap submission with friendly URLs	11	12	13	0	2					
Time Element - 30 days	21	3	1	14	13					
Total Pages indexed (out of 131)	75	26	23	19	17					

Website2 contained 131 dynamic pages. This website had web presence since last one year but 23 webpage(s) of website were indexed in Google. Home page of website was indexed in Yahoo, AOL and Bing. However, website was not index in Ask. After applying several SEO techniques website varied in all search engines. Following graph is descriptive representation of indexing throughout experimentation with step by step application of SEO techniques for website2 (recipe-planner.com).

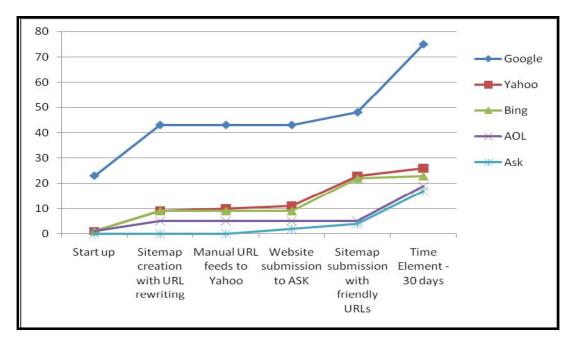


Figure 6.2 Descriptive representation of indexing throughout experimentation with step by step application of SEO techniques for website2 (recipe-planner.com).

6.3.3 Friendly URLs

As discussed in chapter 5 (section 5.2.3), I will be performing some experiment on most controversial variable i.e. user friendly URLs; because I aimed to study how effective this SEO technique is for indexing URLs and how the search engines respond to this technique. This is second part of experiment where I examined if friendly URLs help improve indexing on search engines because unnecessary parameters are not included in the URL. As discussed in literature review, Google claims it does not has any problem reading dynamic looking URLs and it is not required to rewrite URLs to eliminate parameters.

Two samples were taken from the population and URL rewriting (treatment) was applied on experimental group and results were recorded in Excel sheet and basic analysis was performed automatically. Pretest and posttest results are presented in appendix-A. The results are summarized in Table 6.4.

Ta	ıb.	le	6.4	Con	nparison	of ex	perimental	and	control	group	for	friendly	URLs.

recipe-planner.com - sitemap submission with friendly URLs for Experimental Group										
Applied SEO Technique	Experimental Group	Control Group	Experimental Group	Control Group	Difference					
Google	11	12	28.95%	31.58%	2.63%					
Yahoo	12	0	31.58%	0.00%	31.58%					
Bing	11	0	28.95%	0.00%	28.95%					
AOL	1	0	2.63%	0.00%	2.63%					
ASK	1	0	2.63%	0.00%	2.63%					

Table 6.4 contains comparison results of experimental (with friendly URLs as treatment) and control group. Google's response was almost similar towards control and experimental group. Both Yahoo and Bing indexed pages from experimental group without manual submissions of URLs that represent major search engines give preference to friendly URLs.

6.4 Results Analysis

In this section I will interpret the data I collected in the experiment and discuss statistical and empirical analysis performed on collective data. Main objective of this research is to study where search engines have any problem indexing dynamic website or not. The secondary purpose is to test effectiveness of commonly known SEO techniques to help index a dynamic website. Another objective is to investigate common myth of friendly URLs to see if it helps websites getting index or not. To quantify understanding of results (presented in Section 6.3), I created quantitative assessment of effectiveness of each SEO techniques applied on both websites used in research.

To achieve this objective, I organized and summarized data collected during experimentation so that I can analyze and evaluate what I have discovered and present in a form that a conclusion can be derived from it. First, I summarized data and recorded search results taken from search engines and webmaster tools which I used in experiment. This process cost me quite some time because I recorded all data manually as there were not any tools available that I could use to get summary index reports from all search engines. Second, I summarized data in simplest form so that results can be evaluated and described for corresponding SEO technique. Finally, I summarized data so that I can test hypotheses and suggest significant findings.

6.4.1 Outliers

An outlier is a data entry that lies at irregular or far-away distance from other data values in a data set. It is very likely, to see outliers in a study because values do not always follow the normal trend and they appear to be oddly different. However, in this experiment, I did not identify any outlier. Since I was study five different search engines, it was very likely that search results will appear very differently on all of them. Even if I had found such odd entries, I would not consider them as outlier because search engines are not supposed to show similar results for a particular SEO technique.

6.4.2 Efficiency

Since SEO techniques are applied step by step and sample size is also known, it is possible to calculate efficiency of each technique by diving number of indexed webpage(s) after applying each SEO technique to the number of webpage(s) subjected to that SEO technique.

Efficiency = (number of webpage(s) indexed after applying a SEO technique/ total number of webpage(s) exposed to SEO technique) * 100

In Table 6.5 and Table 6.6 efficiency of SEO techniques is calculated for both websites. Since some SEO techniques are effective only for some search engines so N/A signifies that a technique will not affect results of that search engine. In Table 6.5, SEO techniques 6-8 are applied on the experimental group and efficiency is calculated for the each search engine.

Table 6.5 Efficiency of SEO techniques applied on website1 (wasfabththesis.com).

wasfabththesis.com – SEO Techniques' Effectiveness									
Applied SEO Technique	Google	Yahoo	Bing	AOL	Ask				
Startup	N/A	N/A	N/A	N/A	N/A				
Sitemap creation	10%	0%	0%	5%	0%				
Website Submission to SE	N/A	0%	0%	0%	29%				
External Link Building	N/A	5%	5%	10%	0%				
Sitemap submission to SE	29%	0%	0%	17%	0%				
Manual URL feeds to Yahoo	N/A	0%	N/A	N/A	N/A				
Manual URL Submission to Bing	N/A	100%	100%	N/A	N/A				
Internal Link Building on home page	75%	0%	0%	0%	0%				
Time Element - 30 days	22%	21%	21%	7%	7%				

Dynamic URLs were not fully indexed on any of search engines under this study. However, after applying certain SEO techniques on experimental group a significant change was recorded. Manual URL submission to Bing appeared to be most effective SEO technique and Manual URL submission to Yahoo was least effective SEO technique for this website. I found that Yahoo relies on Bing databases to index URLs but not the other way around. When I submitted URLs of experimental group to Yahoo none of them was indexed on Yahoo or Bing. However, in case of URL submission to Bing, results were seen within few hours of submission and URLs of experimental group were indexed on Yahoo as well.

Another SEO technique that helped index webpage(s) in Google is Internal Link Building on Home page. For this SEO technique, I copied some webpage(s) in deeper directories and divided them in two equal groups of (4 dynamic webpage(s)) and linked URLs of experimental group on home pages. Consequently, three out of four webpage(s) were indexed from experimental group and none of the control group webpage(s) were indexed. Index numbers changed on all search engines when this website was left for a period of 30 days.

Table 6.6 Efficiency of SEO techniques applied on website 2 (recipe-planner.com).

recipe-planner.com – SEO Techniques' Effectiveness										
Applied SEO Technique	Google	Yahoo	Bing	AOL	Ask					
Start up	N/A	N/A	N/A	N/A	N/A					
Sitemap creation with URL rewriting	19%	6%	6%	3%	0%					
Manual URL feeds to Yahoo	N/A	17%	N/A	N/A	N/A					
Website submission to ASK	N/A	N/A	N/A	N/A	2%					
Sitemap submission with friendly URLs	30%	32%	34%	0%	5%					
Time Element - 30 days	27%	3%	1%	11%	10%					

All search engines responded differently in terms of indexing URLs for both static and dynamic webpage(s). As presented earlier in the section 6.3, total number of webpage(s) index in the search engines are different and none of the search engines were able to fully index both website completely. It reflects that some SEO techniques are effective for one search engine and some are effective on others. Experimental group, subjected to techniques, responded to some techniques and some other techniques were ineffective. Results will be discussed in the chapter 7 in more details in term of techniques' effectiveness. Figure 6.3 and 6.4 descriptively shows how effective each technique was in getting both websites to get indexed.

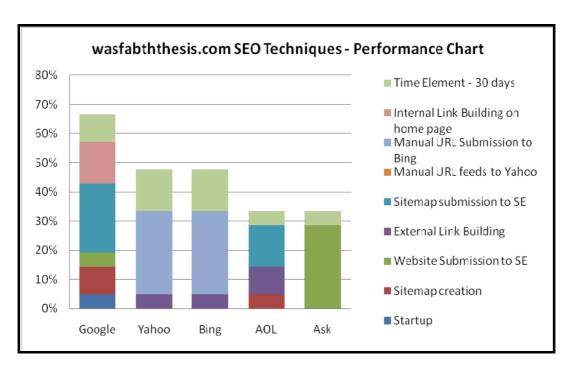


Figure 6.3 Performance chart of SEO techniques of website1 (wasfabththesis.com)

The figure 6.3 shows webpage(s) indexed (percentage) as result of SEO techniques applied on experimental group.

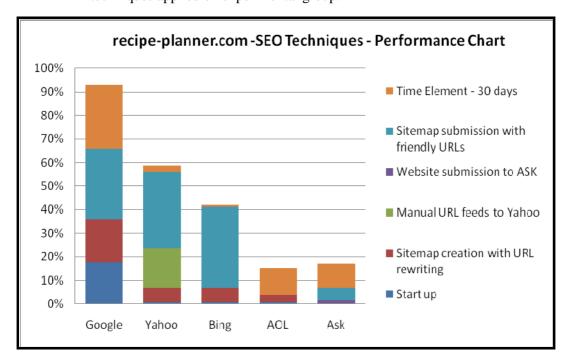


Figure 6.4 Performance chart of SEO techniques for website2 (recipe-planner.com)

The Figure 6.4 shows webpage(s) indexed (percentage) as result of SEO techniques applied on experimental group.

6.4.3 Evaluation of indexing results

One of the objectives to conduct and evaluate indexing results on targeted search engines is to see if search engines have any problem in indexing dynamic website. From the results presented in Table 6.1 and Table 6.3, it is obvious that all search engines have problem in indexing dynamic webpage(s) (either very small like Website1 or relatively large like Website 2) and require additional SEO techniques to optimize them.

After applying SEO techniques presented in Table 6.1, number of webpage(s) indexed on websites gradually increased on all search engines. A significant number of dynamic webpage(s) were not indexed on search engines, including Google, that drives attention to fact search engines don't see dynamic websites very friendly. Static webpage(s) of this website were indexed on all search engines. Figure 6.5 descriptively shows number of indexed and not indexed webpage(s) on website1 (wasfabththesis.com).

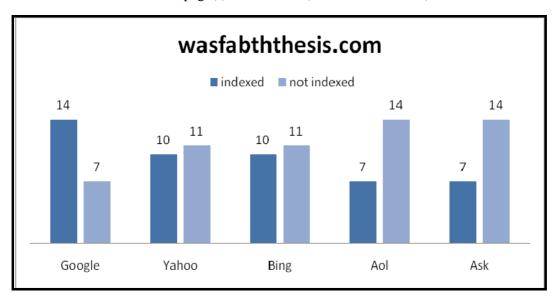


Figure 6.5 Total number of indexed and not indexed static and dynamic webpage(s) on website1 (wasfabththesis.com)

As discussed in chapter 5 (section 5.4), Website2 has web presence since last year but dynamic webpage(s) were indexed only on Google. However, website's home pages were index on Yahoo and Bing as well. After applying SEO techniques, mentioned in Table 6.3, on experimental group more webpage(s) were indexed. The indexed webpage(s) were result of creating friendly URLs for experimental group. I did not performed manual submission of URLS to any of search engines. To demonstrate overall indexing of website2 (recipe-planner.com), results are descriptively presented in Figure 6.6.

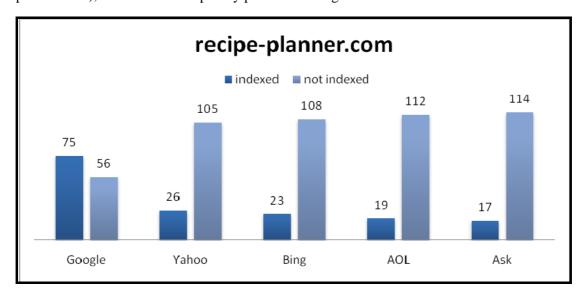


Figure 6.6 Total number of indexed and not indexed dynamic webpage(s) on website2.

As discussed in section 6.3.3 that Results, experimental and control group were taken from the webpage(s) that were not yet indexed on a particular search engine. Indexing results are tested using statistical methods.

6.4.4 Evaluation of friendly URL's results

For this experiment, I aimed to use Website2 (recipe-planner.com) for testing friendly URLs; because it is created by using template engine and it is easier to create friendly URLs. For this purpose I randomly took two sample groups of equal size (38 URLs each), I created friendly URLs on 38 webpage(s) and called them experimental group.

For control group, I did not apply any URL rewriting (treatment) and after creating site maps for both groups I uploaded them on server again. Another reason to use a different website was to control effect of other techniques I applied on Website1. In this experiment, my intention was to see if search engines prefer indexing dynamic webpage(s) with Friendly URLs or prefer traditional dynamic URLs.

As shown in Table 6.4, more dynamic webpage(s) were indexed from experimental group i.e. Google (28.95%), Yahoo (31.58%), Bing (28.58%), AOL (2.63%), and ASK (2.63%) compare to Google (31.58%), Yahoo (0.00%),Bing (0.00%), AOL (0.00%), and ASK (0.00%). The above presented results show that the 28.95% of dynamic webpage(s) are indexed from experimental group. On the other hand, 31.58% webpage(s) were indexed from control group which is more than experimental group. The difference is not such huge; so, it is hard to conclude that Google does not support/promote URL rewriting for making dynamic URLs search engine friendly. But, this figure reflects that Google really made good progress to index dynamic URLs as it used to claim [14]. I will perform statistical analysis to determine significance of results. Results are descriptively represented in Figure 6.7 for comparison purpose.

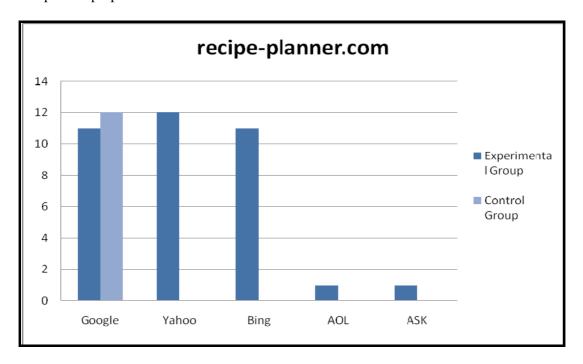


Figure 6.7 Comparison of Friendly URLs between experimental and control group for website (recipe-planner.com).

6.4.5. Normality Testing

Before performing any statistical tests that assume normality of data, it is necessary to do normality test on the data [47] [48]. I performed normality test of data collected for both

websites. Generally data is of two kinds: parametric and non-parametric. Data is called parametric if it shows normal distribution of curve, otherwise data is called non-parametric [47] [48]. Wohlin *et al.* [43] discussed that if experimenter is not sure about nature of data then non parametric tests should be used because parametric tests are considered to be more powerful for than non-parametric tests. I performed Shapiro-Wilks Normality Test on data to determine nature of sample data.

6.4.6. Shapiro-Wilks Normality Test

This test is used to determine if sample data is normally distributed or not [49]. This test rejects the null hypothesis when the Probability (P) value of W is very small. The W is measure of the straightness of the normal probability plot. It is common belief among researcher that for small to medium data Shapiro-Wilks normality test is most reliable to verify data distribution [49] [50].

I used stats-direct software tool to calculate Shapiro-Wilks test. I tested sample data for both websites using this tool. Results of Shapiro-Wilks normality test are presented in Appendix-B.

6.4.7. Non-Parametric Test

Shaprio-Wilks normality test results showed that data is not normally distributed for data collected for indexing and friendly URLs. As discussed to reject Null hypotheses (H_{01} and H_{02}) I need to perform non-parametric tests. First of all I will perform testing for H_{01} , to test if targeted search engines (Google, Yahoo, Bing, AOL, ASK) have any problem indexing dynamic websites. If H_{01} will be rejected, I will proceed to perform this test on data collected for friendly URLs experiment. Since I am studying effects of friendly URLs as treatment against control group, I chose to use Mann-Whitney Test which works best for both tests.

6.4.8. Mann-Whitney Test

This test is non-parametric test and used to compare two independent groups of sample data [51]. This non-parametric test is based on assumption that sample under consideration are drawn from same population so difference between average values will be zero or both sample have same median value[51] [52]. Results of Mann-Whitney test will be discussed in next section during hypotheses testing. Details Results of Mann-Whitney test are provided in Appendix-B.

6.5 Hypotheses Testing

I stated two Null hypotheses:

Hoi: Dynamic websites do not require search engine optimization (SEO) since three major search engines are equipped to index them as it.

*H*₀₂: *SEO* techniques do not make significant difference in indexing dynamic websites on three major search engines.

It is obvious from data (Table 6.1 and Table 6.3) presented in section 6.3 that none of the search engines are able to completely index any of two websites under experimentation. The application of Mann-Whitney test (non-parametric test) has also provided the evidence that results are significant because value of p is less than level of significance (0.05) for data summarized for indexed webpage(s) on both websites. So results are significant for both

websites and we can reject Null hypothesis H_{01} on the basis of results and evidence given in Appendix-B

Since we rejected Null hypothesis H_{01} , I performed Mann-Whitney Test on data summarized for friendly URLs presented in Table 6.4 (Section 6.3.3) to test second Null hypothesis H_{02} . In this test, I found that value of p is less than level of significance (0.05). So, results are significant that we can reject Null hypothesis H_{02} on the basis of results give in Appendix-B.

Based on statistical tests results, I rejected both Null hypotheses (H_{01} and H_{02}) that would mean that search engines cannot index dynamic websites on their own and they require additional SEO techniques to help dynamic URLs index on search engines.

7 DISCUSSION

Main purpose of this section is to share my experiences obtained from research, literature survey and empirical experiments. Based on my finding, I will answer research questions addressed in this study. I will also discuss lesson learned while performing this course of study.

E-commerce has drastically evolved in the past decade with changing consumer shopping behavior. As result of this trend, many businesses are emerging with e-commerce and e-marketing. E-commerce is continually evolving with fundamental changes in technology. Since search engines are key source of locating reliable businesses, businesses appearing in top search results have substantially competitive advantage over those that don't. E-commerce websites are mostly dynamic in nature and contents change on daily basis, so such websites are harder to optimize. It is possible to promote a business by paying to search engines to appear in top search results. However, for businesses who want to appear in generic search result could endorse their websites by SEO techniques. Therefore SEO has become an essential part of web development process. There are many SEO companies around the globe and this number is continually increasing. I found this topic interesting because search engines are continually evolving with enhance indexing capabilities; and I wanted to study significance of SEO techniques for dynamic websites and behavior of major search engines on them.

Before starting literature review, I decided to perform preliminary research on dynamic websites. I read numerous online articles, forums and blogs about controversial topics about dynamic websites such as friendly URLs, URL depth, and URL parameters. It was obvious from my initial research that there are many misconceptions about indexing of dynamic websites. Moreover, there was not much information available about search engines other than Google. Therefore, I performed detailed literature survey on dynamic websites and SEO techniques that can improve their visibility in targeted search engines (Google, Yahoo and Bing).

7.1 Literature Review

In the following proceeding section, I will present answer of RQ1 based on the detailed literature survey that I carried out in my research work. Also I would discuss what I concluded from literature review.

On-page SEO techniques: In SEO campaign, On-page SEO starts right from the birth of website almost carry out throughout the life the website. Most of elements On-page SEO improve ranking of website; however, some elements are important for indexing point of view. Title tag and Meta tags are extremely important elements of on-page SEO techniques for search engines as well as for users. As every search engines use contents of Title and Description Meta tag to display in search results which ultimately boost CTR of website. It is obvious that some search engines do not give such importance to Meta tags for ranking of website. However, these tags are almost equally important for indexing in all search engines. The important Meta tags include Robot Meta tag, description Meta tag and keyword Meta tag. Robot Meta tag use explicitly to direct search engines that which webpage(s) are required/not required indexing or following them. Expiry Meta tag is important for rapidly updating websites to refresh and get space in search engine indexes for fresh contents. Bread-crumb trail tool provides an additional text-based navigational way through internal

Bread-crumb trail tool provides an additional text-based navigational way through interna webpage(s). It reduces chances of search engine to miss indexing of important webpage(s).

URL size and structure another important element of On-page SEO that need a good attention for indexing. Domain name of website should decide carefully according aim of website. Moreover, long and complicated URLs with a lot of parameters and special

characters are problematic for search engines. Every search engines recommends avoiding such URLs. Use of ALT tag makes images information understandable for search engines; moreover, it allows images of website searchable through search engine.

Off-page SEO techniques: These techniques are thought as to implement for better ranking of website but some of them are important even for indexing point of view. Some of significant off-page SEO techniques include directory submission, anchor Text and external link building and creation of forums and blogs. These factors allow targeted website to get in eye of search engines from other sources i.e. directories and other linking websites. The creation of forums and blogs allow website's users to add comments and to read blogs that ultimately increase traffic and keep website up-to-date. Additionally, these factors attract search engines to come back to target website for crawling and indexing freshly added content on webpage(s).

Design considerations for friendly websites: Frames are used to facilitate site user for easy access to the contents of website; but search engines cannot follow frames contents. Therefore, they should be avoided or contents of frames should provided by alternative way i.e. <noframe> tag. Sometimes dynamic websites need to use forms to get input from user. But these are also troubling contents for search engine. Therefore, they should optimize too. JavaScript and Flash are used to make website attractive for user point of view. However, search engines completely avoid JavaScript to crawl; so it should be avoid or need to optimize them. Unnecessary use of Flash increase page size that is not appealing for user as well as for search engines. As both of these contents are not indexed by search engines therefore Flash contents should be avoided or optimized.

Search engines essential files for SEO: Sitemap (XML site map) and robot.txt files are needed explicitly to search engines on target website. The XML Sitemap is accepted as a standard by major search engines (Google, Yahoo, Bing, Ask and AOL). It is provided with a list of important webpage(s) of website that cannot be discovered directly through website hierarchy. Robot.txt is explicitly used to direct search engines towards webpage(s) are needed/or not needed to be indexed /followed. It saves crawling time of search engine by stopping it to crawl or index unnecessary webpage(s).

URL rewriting and redirecting: URL directing becomes important when contents of indexed website are likely moved new other location (domain and database). Redirecting provide direct search engine to new URL (location) without page not found error which not only loose traffic of website also it discourage search engine come back to website.

Search engine (including Yahoo, Bing, AOL and Ask) recommend using simple and static URLs. Rewriting provide a way to cleanup complex dynamic URLs to simple friendly URLs. Google does not recommend rewriting of dynamic URLs to friendly URL. On the other hand, other search engines (Yahoo and Bing) prefer friendly and static URLs and recommend avoiding complex/dynamic URLs. If it is needed to optimize a dynamic website on search engines other than Google then comprise with dynamic URLs could be quite paying. Although Google does not recommend using URLs encoding on dynamic URLs, top search results of Google are mostly URLs containing search phrases or keywords in them. That clearly indicates Google's crawler rate URLs containing search words in them better than dynamic looking URLs.

Major search engine's guidelines for SEO: Google, Yahoo and Bing provide some guidelines for helping out web developers for healthier indexing of websites. Nevertheless, they do not provide any hard and fast rules and instructions to support ranking. Still these guidelines assist a lot to optimize websites for better indexing. Search engines have almost same criteria for indexing. There is no such a strict contradiction among their indexing criteria

Useful search engine's tools for SEO campaign: Major search engines (including Google, Yahoo and Bing) provide tools for SEO campaign on their official websites. These tools are free of cost and very helpful for tracking indexing and ranking of websites on specific search engine. Moreover, some tools provide details if there is problem related to indexing or crawling of website(s). These tools comprise Google webmaster tool, Yahoo site explorer and Bing Webmaster tool provide facilities like submission of website, submission of

Sitemaps, manual submission of URLs of website, automatic Sitemap generation, removal of indexed webpage(s) and to track detail information that why any webpage(s) are not be indexed, on Google, Yahoo and Bing respectively. Google Analytics is one of most useful tool to track indexing and ranking of websites on Google. Moreover, YSlow for Firefox and Webrank Toolbar are Firefox add-on. The former helps to provide detail that why website is slow also assist how to accommodate and optimize size of webpage(s). The latter provide ranking and indexing details on several search engines.

The answer of RQ1 can be summarized as:

The current state of art of SEO for dynamic website can be reflected by increasingly number of dynamic website for business and their appearance in search engines. As, in early 2000's search engine were not able to index most of website's contents including frames, forms, images and dynamic websites. Today, the element of SEO i.e. On-page SEO techniques and Off-page SEO techniques are considered as heart of SEO for static as well as for dynamic websites. These SEO techniques made most contents of websites index-able by most of search engines. SEO techniques i.e. rewriting and redirecting have been highly adopted by website developers to overcome indexing problem of dynamic URLs.

Google, Yahoo and Bing are providing design and technical guidelines for web developers for optimizing dynamic website. Along with that search engines allow to use their tool to get insight of indexing related issues and problems. These tools are obliging for promoting SEO campaign for static as well as for dynamic websites.

7.2 Empirical Study

In the following proceeding section, I will present answer of RQ2 based on the results derived from experimental research I carried out on two sample websites.

Search Engines and indexing dynamic websites: All search engines response differently towards dynamic websites. None of the targeted search engines was able to fully index the dynamic websites used in the experiments. I used two websites having different sizes and ages and none of them was fully indexed. It is interesting to notice that Google crawlers are more efficient than other search engines. If a dynamic website has a sitemap, and it is manually submitted to Google then it is very likely that significant number of webpage(s) will be indexed without performing any other SEO technique. Home page link building is also an efficient and supportive technique that helps webpage(s) indexed on Google. It is interesting to notice that Google does not have a fundamental problem indexing dynamic looking webpage(s).

Bing and Yahoo require SEO to index dynamic websites. Both search engines allow to feed dynamic URLs with parameters. However, manual URLs submission to Yahoo is not efficient at all because Yahoo does not guarantee indexing submitted URLs in Yahoo Site Explorer. Conversely, URL submission to Bing is very efficient not only because all submitted URLs are indexed in are timely manner, but URLs fed in Bing Webmaster Tool get indexed in Yahoo search engine as well. In addition to that, it provides search engine experiments and allows downloading website index and crawling related report.

It seems like a quick and easy way to index dynamic webpage(s) on two search engines without getting into hassle of lingering SEO techniques such as friendly URLs. It might be a good solution for small website but not for large website where numbers of webpage(s) are growing every day.

It is not possible to submit all new URLs to Bing on daily basis for two reasons. First, there is a limited monthly quota (50 URLs per month for Bing) of URLs submission in Bing and Yahoo. Second, website owner needs to manually submit URLs to search engine's webmaster tool. Again for large websites containing thousands of webpage(s) it is much a hassle. This technique has minimal effectiveness. So, there is need to have SEO technique(s) which help dynamic webpage(s) index automatically without going through URL submission process. This is what I evaluated by writing friendly URLs for one website and results

showed that these webpage(s) were indexed on Google, Bing, and Yahoo without manual submission.

Friendly URLs: As discussed earlier in literature review that Google claims it has no problem manipulating dynamic URLs. However, it is also observed in this study that Google index about 57% URLs with about equal number of dynamic and static looking URLs from the experimental and control group of website2. So, there is need to do addition SEO techniques to help index remaining webpage(s).

Friendly URLs have an advantage of dynamic looking URLs that encoded URLs are more likely to index in Google, Yahoo, and Bing without manually submitting URLs to search engines. So, friendly URLs should be preferred over dynamic looking URLs.

Other Effective SEO Techniques: All SEO techniques, either specific to dynamic website or common to both static and dynamic website, have different impressions on search engines (Google, Yahoo, Bing, Ask, and AOL). However, some techniques improve results on all search engines such as sitemap submission and encoded URLs.

To keep scope of this thesis limited, I used a few SEO techniques that are helpful in indexing a dynamic website. I found that sitemap creation is one of few efficient techniques which are helpful for all search engines. When a dynamic website is created with friendly URLs and sitemap is submitted to search engines, I found significantly improved indexing results. This technique is effective for Google, Yahoo and Bing. Ask and AOL relies on sitemap as well to index a website.

All search engines consistently update their index databases to include add/remove URLs. So, a website that has web existence for some times is more likely to have it webpage(s) in index databases of search engines. As observed in this experimental research, result varied on all search engines during the wait period (about 30 days). In addition to these techniques, linking new URLs on indexed webpage(s), particularly home page, of a site also helps new webpage(s) to get indexed.

The experiments result reflects some facts that Ask Search prefers dynamic websites that do not have URLs written for them. Both Yahoo and Bing prefer friendly URLs over typical URLs with parameters. Although search engines crawlers are able to see dynamic pages of a website, none of the targeted search engines was able to fully index dynamic website without SEO techniques. The experiment results showed that number of dynamic webpage(s), with encoded URLs, indexed in Google (28.95%) is less than compared to Yahoo (31.58%) and Bing (28.95%), URL encoding is highly recommend not only because encoded webpage(s) are more likely to indexed in Bing and Yahoo but also because friendly URLs are more likely to appear in top results in Google.

The answer of RQ2 can be summarized as:

The current state of the art SEO is relatively straightforward process for static webpage(s); however, the process is much more complex for dynamic websites and requires SEO grounding before website development starts. The idea that search engines can only index webpage(s) that actually exist is not obsolete because all major search engines require suitable SEO techniques to index dynamic websites. Google crawlers are much more competent but not outstanding, in indexing dynamic webpage(s) than Yahoo, Bing, Ask and AOL. From experiment results, it is obvious that with appropriate SEO techniques it is possible to improve indexing of dynamic websites to a good extent on all major search engines.

8 CONCLUSION AND FUTURE WORK

In conclusion, I will summarize the steps I went through during the research process. I will also address problem I ran into and lesson learned from this research work. In addition to that, I will discuss whether I was able to commence aims of this research work or not. Finally, I will summarize my overall findings and answer research questions. Further in future work, I will discuss future work that I would like to perform in the area of SEO for dynamic websites.

8.1 Conclusion

SEO is essential for websites to increase their visibility in search engines. On-page SEO techniques and Off-page SEO techniques are commonly used SEO techniques for websites. On-page SEO techniques help in optimizing website contents by making it accessible/understandable for search engines; because search engines can understand only certain formats of website contents. Whereas, Off-page SEO is nothing to do with website contents but it affects websites to get better ranking and improve crawlers' visit ratio to the websites. There are many useful SEO tools available for webmasters such as for Google: Google Webmaster tools, Google Analytics, and Google Adword (paid advertisement); for Yahoo: Yahoo Site Explorer; for Bing: Bing Webmaster Tool. Among these tools Bing Webmaster tool and Google Analytics are most effective for SEO point of view. Among browser extensions (add-ons), YSlow is very useful tool which provides a simulated environment to view changes in a page along with page content information. Since different search engines work with different algorithms to index dynamic websites, search results are different in search engines.

Major search engines like (Google, Yahoo and Bing) claims that there is no problem to index/ rank static websites (if they are optimize); however, there are some conflicting facts about indexing of dynamic websites. Search engines are continually evolving with ability to index dynamic URLs (particularly Google), experimental research of this study explored to what extent search engines have progressed. To be specific, I aimed to study indexing behavior of Google. Yahoo, Bing, Ask and AOL. Results showed that none of targeted search engines are able to fully index dynamic websites. I applied some SEO techniques to help indexing dynamic webpage(s) and tried to determine to what extent dynamic websites can be optimizes. A technique like URL rewriting can help a dynamic website keep its URLs constant, so that changing product code and location does not affect URLs. Because in standard dynamic URL contain category and products code as part of the URL; if product is moved in the table it's URL will also change. Moreover, dynamic webpage(s) with friendly URLs are more likely to index on search engines. Based on experiment results and analysis, I concluded that SEO techniques are essential to index dynamic websites.

The experimental results showed that URL encoding of dynamic webpage(s) improved indexing results by 31.58% and 28.95% for Yahoo and Bing respectively. These are significant figures to reflect the effectiveness of URL encoding URL/friendly URLs on these search engines. Conversely, Google result does not reflect such positive (though not negative) response towards friendly URLs. The URL encoding is highly recommend not only because URL encoded webpage(s) are more likely to indexed in Bing and Yahoo but also because friendly URLs are more likely to appear in top results in Google.

Since a properly indexed websites are an initial step towards better ranking to bring websites more visible in search engines, it is important to optimize a websites properly to increase credibility of a websites. It is possible to index dynamic webpage(s) on all search engines with favorable SEO techniques. Not one technique works on all search engines so it is necessary to deem individual consideration of all search engines.

8.2 Future work

During the course of this research work, I found some interesting issues; which were not covered in this study due to time constraints and limited scope of this research work.

Webpage(s) from website bearing higher rank (credibility) are more likely to appear among top search results of a search engine. Main aim of SEO techniques is to improve overall visibility in search engines and help it obtain higher rank. However, the scope of my research was limited to explore SEO techniques that help to improve indexing of dynamically generated webpage(s). In most cases, ranking takes more than three months to show clear results of how successful the efforts are for improving website ranking. Due to time and scope constraints I was not able to perform any experiments on techniques that can help improve website ranking. So in future, I would like to perform a detailed study on techniques that can help achieve a real website acquire higher ranking. I would also like to explore that whether it is possible to achieve top/better ranking in Google, Yahoo and Bing at the same time by using some common SEO techniques.

Success of an online business directly depends on consumer's shopping experience so websites administrator and writer want to keep track of buyer's trends. I would also like to do research on optimization based on online trends and consumer behaviors.

REFERENCES

- [1] X. Zhang, M. Zuo, & Q. Liu;, "Analysis of the Reasons Why Invisible Web Can't Be Seen and its Effective Retrieval Strategies", *Innovative Computing Information and Control (ICICIC) '08 3rd International Conference on*, vol., no., pp.563-563, 18-20, (June 2008)
- [2] M. Alfano & B. Lenzitti;, "A Web Search Methodology for Different User Typologies" *International Conference on Computer Systems and Tech----nologies-CompSysTech*" (2009)
- [3] C. Sheram;, "Google Power: Unleash the full potential of Google", *McGraws-Hill/Osborne Companies*, *ISBN 0-07-225787-3*, (2005)
- [4] Y. Ru & E. Horowitz;, "Indexing the invisible web: a survey", *Department of Computer Science, University of Southern California, Los Angeles, California, USA page* 2, (2005)
- [5] L. Huang;, "Challenging the Invisible Web: Improving Web Meta-Search by Combining Constraint-based Query Translation and Adaptive User Interface Construction", The Department of Computer Science of Technische Universität Darmstadt, Germany, (2003)
- [6] J. Köhne;, "Optimizing a large dynamically generated website for search engine crawling and ranking", Department of Media- and Knowledge Engineering Technical University of Delft Netherlands, (2006)
- [7] E. Enge, S. Spencer, R. Fishkin & J. C. Stricchiola;, "The Art of SEO: Mastering Search Engine Optimization", O'REILLY Media Inc., ISBN 978-0-596-51886-8, (October 2009)
- [8] G. Rogan;, "A study of Search Engine Optimization methods", National University of Ireland, Gateway (2009).
- [9] N. Nazar;, "Exploring SEO Techniques for Web 2.0 Websites", Master of science Thesis in software Engineering and Technology, Department of CS and Engineering Chalmers University Of Technology Göteborg, Sweden, (June 2009)
- [10] A. Pirkola;, "The effectiveness of Web search engines to index new sites from different countries", *Information Research*, *Department of Information Studies*, *University of Tampere*, 33014 Tampere, Finland, (2009)
- [11] R. Vadivel & Dr. K. Baskaran;, "Implementing Search Engine Optimization Technique to Dynamic/Model View Controller Web Application", *Global Journal of the Computer Science and Technology Vol. 10 Issue 6, V1.0, (July 2010)*
- [12] C Duda, G Frey, D. Kossmann & C. Zhau;, "AJAXSearch: Crwaling, Indexing & Searching Web 2.0 Applications", Prfocedding of the VLDB Endowment v.1 n.2, (2008)
- [13] J. L. Ledford; "SEO: Search Engine Optimization Bible", *John Wiley & Sons, Inc. Publishing USA, ISBN 0470452641, (2009)*
- [14] Official News on Crawling and Indexing sites for Google index.

 http://googlewebmastercentral.blogspot.com/2008/09/dynamic-urls-vs-static-urls.html

 (Accessed on 2010-08-11)
- [15] How Dynamic Pages Are Indexed in Yahoo!

- http://help.yahoo.com/l/us/yahoo/search/indexing/indexing-05.html;_ylt=AtHd2LCQTV20maZncciizPZYqCN4 (Accessed 2010-08-25)
- [16] Bing Webmaster Center FAQs

 http://www.microsoft.com/downloads/en/confirmation.aspx?displaylang=
 en&FamilyID=05bcf5b6-c0ea-4bef-b381-72434e826681
 (Accessed on 2010-08-25)
- [17] comScore, Inc. Releases November 2010 Search Engine Rankings
 http://www.comscore.com/Press_Events/Press_Releases/2010/12/comScore_Releases_November_2010_U.S. Search_Engine_Rankings
 (Accessed on 2010-09-12)
- [18] Dag I. K. Sjøberg, T. Dybå & M. Jørgensen; "The Future of Empirical Methods in Software Engineering Research", *Simula Research Laboratory, Lysaker, Norway* (IEEE) ISBN: 0769528295 (2007).
- [19] C. Hart;, "Doing a Literature Review: Releasing the social science research imagination", Sage Publication London, ISBN: 0761959742 (2002)
- [20] S. Odom & S. R. Odom;, "SEO for 2010: Search engine Optimization Secrets", MediaWorks Publishing Inc, USA, ISBN: 978-0-557-16133-1, (2010)
- [21] M. H. Fleischner;, "SEO Made Simple: Strategies for Dominating the World's Largest Search Engine" published by Lightning Source United States of America, ISBN13: 978-0-615-17863-9, (2008)
- [22] "Google Search Engine Optimization Starter Guide", Google webmaster central (2010).
- [23] D. Viney; "Get to the Top on Google" *Nicholas Breealey Publishing London ISBN:* 978-1-85788-502-6, (2008)
- [24] J. Carpenter;, "Steps for Dynamic Site Search Engine Optimization", *Mea digital White Paper*, (2006)
- [25] S. M. Spencer;, "Search engine optimization Best and Worst Practices", netconcepts white paper, (2008)
- [26] S. Wiideman;, "SEO in a Day", http://www.top10seotips.com/pdf/SEO-in-a-Day.pdf (Accessed on 2010-10-20)
- [27] Sitemaps XML format http://www.sitemaps.org/protocol.php (Accessed on 2010-12-21)
- [28] Frames and search engines, http://searchenginewatch.com/2167901 (Accessed on 2010-12-22)
- [29] Google Webmaster Central "First click Free"

 http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer="74536"
 (Accessed on 2010-12-23)
- [30] Google Webmaster Guidelines
 <a href="http://www.google.com/support/webmasters/bin/answer.py?answer="http://www.google.com/support/webmasters/bin/answer
- [31] S. Shivakumar; (Google Engineering Director) "Webmaster-friendly", http://googleblog.blogspot.com/2005/06/webmaster-friendly.html (2005)
- [32] Yahoo Guidelines about Search Indexing and Ranking http://help.yahoo.com/l/us/yahoo/search/indexing/; ylt=AgDalMQ. BPkES5a9.6oh0qlYqCN4 (Accessed on 2010-11-11)
- [33] Bing Guidelines for Search Engine Optimization
 http://help.live.com/help.aspx?mkt=en-US&project=wl_webmasters
 (Accessed on 2010-12-19)
- [34] Google Webmaster Tools

- http://www.google.com/support/webmasters/?hl=en (Accessed on 2010-12-19)
- [35] Google Analytics http://www.google.com/analytics/ (Accessed on 2010-12-10)
- [36] Google AdWords
 http://adwords.google.com/support/aw/?hl=en (Accessed on 2010-12-04)
- [37] Yahoo! Site Explorer
- http://help.yahoo.com/l/us/yahoo/search/siteexplorer/ (Accessed on 2010-12-07)
- [38] Bing Webmaster Tools http://www.bing.com/toolbox/webmasters/ (Accessed on 2010-12-16)
- [39] "Firefox (Addon) WebRank Tool for Google, Yahoo and Bing", http://www.deerawan.com/software/firefox-addon-webrank-toolbar/ (Accessed on 2010-12-23)
- [40] S. Brin & L. Page; "The Anatomy of a Large-Scale Hypertextual Web Search Engine", Computer Science Department, Stanford University, Stanford, CA 94305, USA, (1998)
- [41] "The History of Yahoo! How It All Started", Yahoo! Inc http://docs.yahoo.com/info/misc/history.html (Accessed on 2010-11-23)
- [42] A. Wall; "Search Engine History" http://www.searchenginehistory.com/#microsoft (Accessed on 2010-12-18)
- [43] C. Wohlin, P. Runeson, M. Ho. st, M. Ohlsson, B. Regnell & A. Wessle'n;, "Experimentation in Software Engineering An Introduction", *Kluwer Academic Publishers*, (2000).
- [44] T.D. Cook & D.T. Campbell; "Quasi-experimentation. Design and analysis issues for field settings", *Chicago: Rand McNally* (1979).
- [45] D. T. Campbell & J. C. Stanley;, "Experimental and quasi-experimental design for Research", American Educational Research Association", Chicago: Rand McNally (1963).
- [46] R. C. Martella, R. Nelson, & N. E. Marchand-Martella;," Research methods: Learning to become a critical research consumer", *Boston: Allyn & Bacon* (1999).
- [47] "Wolfram MathWorld Normal Distribution", http://mathworld.wolfram.com/NormalDistribution.html (Accessed 2010-12-27).
- [48] M. Hollander & D. A. Wolfe;, "Nonparametric Statistical Methods (2nd.)", Wiley-Interscience, (1999).
- [49] S. S. Shapiro &, M. B. Wilk;, "An analysis of variance test for normality (complete samples)", Biometrika, 52, 3 and 4, pages 591-611(1965).
- [50] E. M. Saniga & J. A. Miles, "Power of Some Standard Goodness-of-Fit Tests of Normality against Asymmetric Stable Alternatives", *Journal of the American Statistical Association, Vol. 74*, pp. 861-865 (1979).
- [51] H. B. Mann & D. R. Whitney;, "On a test of whether one of 2 random variables is stochastically larger than the other", *Annals of Mathematical Statistics*, 8,50-60 (1947).
- [52] "Wolfram MathWorld Mann-Whitney Test Statistic", http://mathworld.wolfram.com/Mann-WhitneyTestStatistic.html, (Accessed 2010-12-28)

APPENDIX-A

PRETEST POSTTEST SCREENSHOTS

The Appendix A is presented with Pretest results i.e. before applying SEO technique (Friendly URL) and Posttest results i.e. after applying SEO technique (Friendly URL) of website2 on Google, Yahoo, Bing, AOL and Ask search engines.

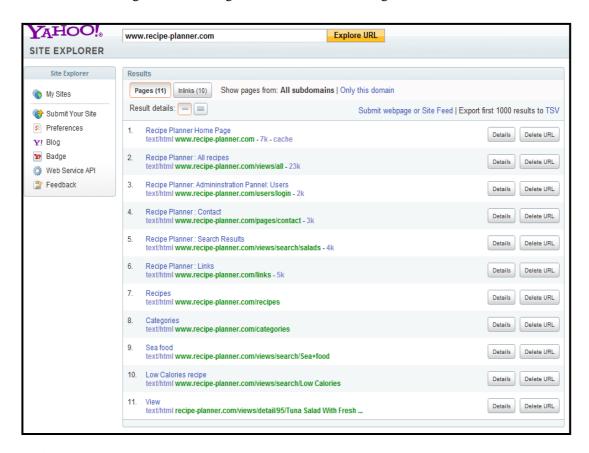
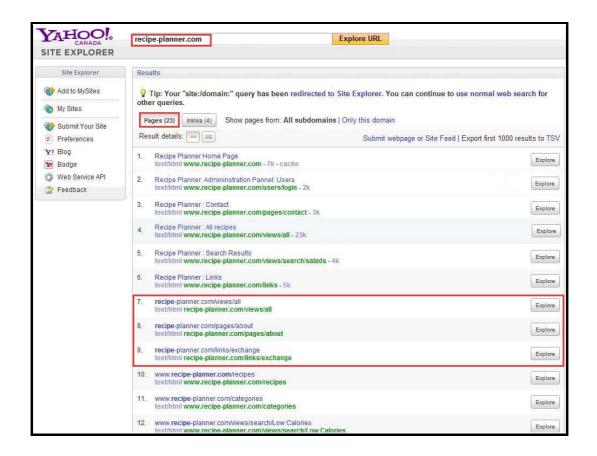


Figure 5.1 Pretest on Yahoo: Sitemap submission with friendly URLs.

The screenshot presented above is Pretest results i.e. before applying SEO technique (Friendly URL) of website2 on Yahoo search engine.



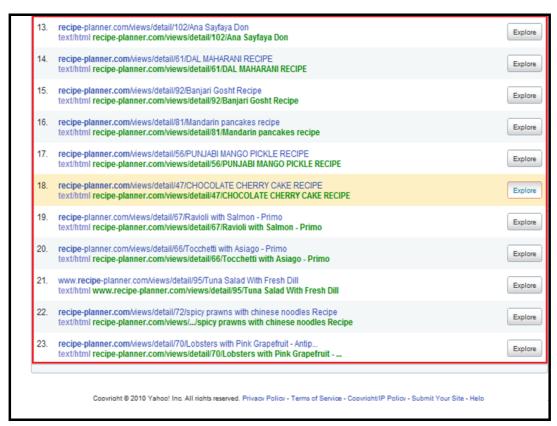


Figure 5.2 Posttest on Yahoo: Sitemap submission with friendly URLs.

The screenshot presented above is Posttest results i.e. after applying SEO technique (Friendly URL) of website2 on Yahoo search engine.

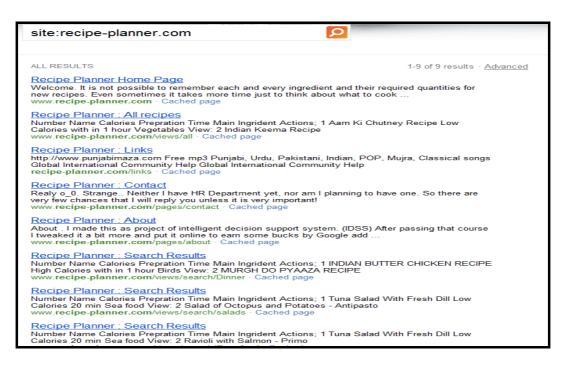


Figure 5.3 Pretest on Bing: Sitemap submission with friendly URLs.

The screenshot presented above is Pretest results i.e. before applying SEO technique (Friendly URL) of website2 on Bing search engine.



Recipe Planner Home Page

Welcome. It is not possible to remember each and every ingredient and their required quantities for new recipes. Even sometimes it takes more time just to think about what to cook ... www.recipe-planner.com · Cached page

Recipe Planner: Welcome

Abstract We are aim to develop a system to support user to get recipes according to their provided criteria, taste and interest. This report explains briefly steps for our ... www.recipe-planner.com/pages/welcome · Cached page

Recipe Planner: - Bresaola with Flakes of Parmesan - Appetizer

Bresaola with Flakes of Parmesan - Appetizer. I show here a simple and fast way to prepare a delicious Italian appetizer made with a typical Italian ingredient that is bresaola of ... recipe-planner.com/views/detail/89/Bresaola%20with%20Flakes%20of%20Parmesan%20-%20... · Cached page

Recipe Planner : - NIMBU KA MITHA ACHAR (Sweet Lemon pickle) RECIPE

NIMBU KA MITHA ACHAR (Sweet Lemon pickle) RECIPE. Ingredients: 340gms smooth-skinned Lemons 2tbsp salt 1/2tsp Fennel seeds (crushed) 1/2tsp Cumin seeds (crushed) recipe-

planner.com/views/detail/55/NIMBU%20KA%20MITHA%20ACHAR%20(Sweet%20Lemon%20... Cached page

Recipe Planner: Problem Solution

Proposed System: Our system is supposed to help those to resolve their problem that what to cook and how to cook just by providing some criteria.

www.recipe-planner.com/pages/problem_solution · Cached page

Recipe Planner: Links

http://www.punjabimaza.com Free mp3 Punjabi, Urdu, Pakistani, Indian, POP, Mujra, Classical songs Global International Community Help Global International Community Help recipe-planner.com/links · Cached page

Recipe Planner: - spicy prawns with chinese noodles Recipe

spicy prawns with chinese noodles Recipe. Ingredients 350g.chinese egg noodles 6 medium tomatoes 1-2 fresh red chillies 6 spring onions 2-3 cloves 2 in. fresh root ginger recipe-planner.com/views/detail/72 · Cached page

Recipe Planner: - Jugu Cake / Twice Baked Peanut Bars

Jugu Cake / Twice Baked Peanut Bars. Jugu or njugu is peanut or groundnuts in Swahili. As a child I remember them as karanga which is also swahili. recipe-planner.com/views/detail/107 · Cached page

Recipe Planner: Search Results

Number Name Calories Prepration Time Main Ingrident Actions; 1 Tuna Salad With Fresh Dill Low Calories 20 min Sea food View: 2 Ravioli with Salmon - Primo www.recipe-planner.com/views/search/Sea+food · Cached page

Recipe Planner: - CHOCOLATE ANGEL CAKE

CHOCOLATE ANGEL CAKE. Ingredients: 1/4 cup Cocoa Powder 1/4 cup Hot Water 1 1/2 cup Granulated Sugar 3/4 cup Fine Wheat Flour (Maida) 1/4 tsp Salt 12 Eggs White recipe-planner.com/views/detail/48/CHOCOLATE%20ANGEL%20CAKE · Cached page



Figure 5.4 Posttest on Bing: Sitemap submission with friendly URLs.

The screenshot presented above is Posttest results i.e. after applying SEO technique (Friendly URL) of website2 on Yahoo search engine.

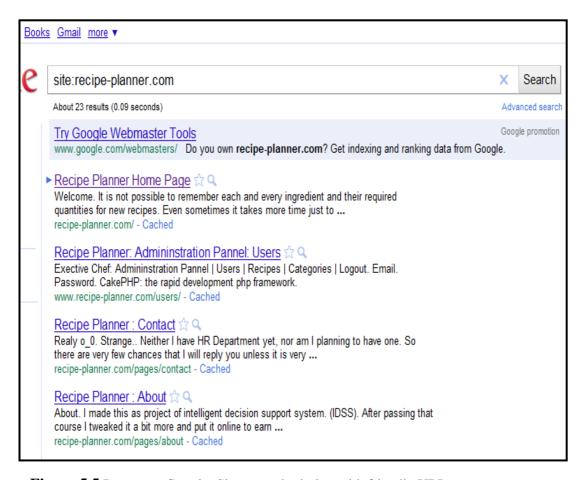
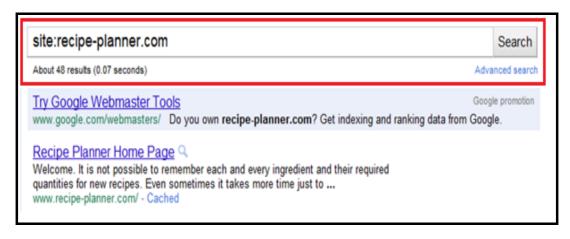


Figure 5.5 Pretest on Google: Sitemap submission with friendly URLs.

The screenshot presented above is Pretest results i.e. before applying SEO technique (Friendly URL) of website2 on Google search engine.



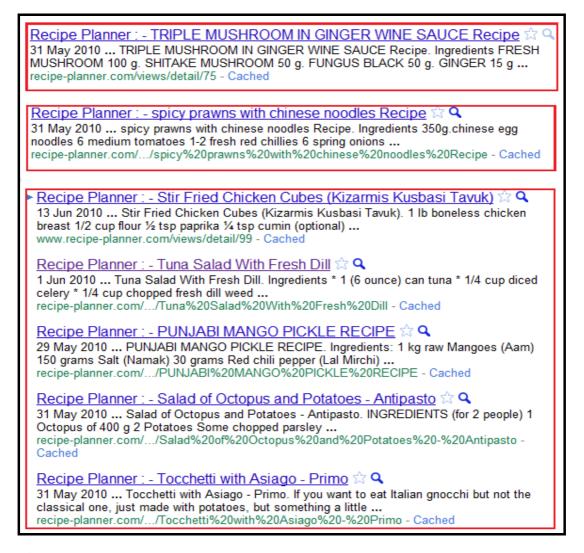


Figure 5.6 Posttest on Google: Sitemap submission with friendly URLs.

The screenshot presented above is Posttest results i.e. after applying SEO technique (Friendly URL) of website2 on Google search engine.

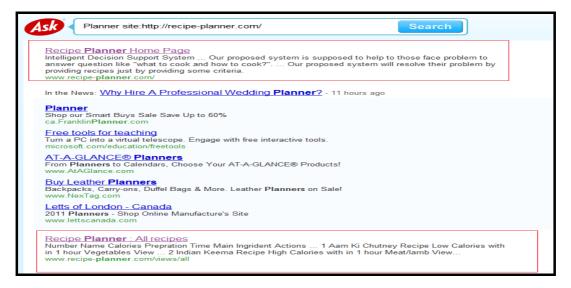


Figure 5.7 Pretest on Ask: Sitemap submission with friendly URLs.

The screenshot presented above is Pretest results i.e. before applying SEO technique (Friendly URL) of website2 on Ask search engine.

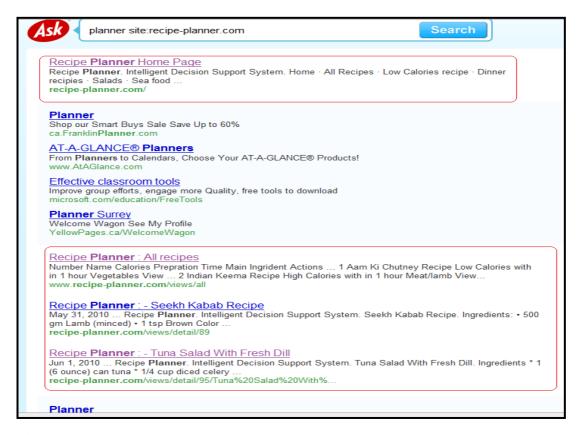


Figure 5.8 Posttest on Ask: Sitemap submission with friendly URLs.

The screenshot presented above is Posttest results i.e. after applying SEO technique (Friendly URL) of website2 on Ask search engine.

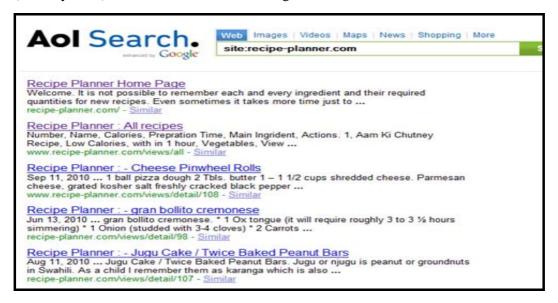


Figure 5.9 Pretest on AOL: Sitemap submission with friendly URLs.

The screenshot presented above is Pretest results i.e. before applying SEO technique (Friendly URL) of website2 on AOL search engine.

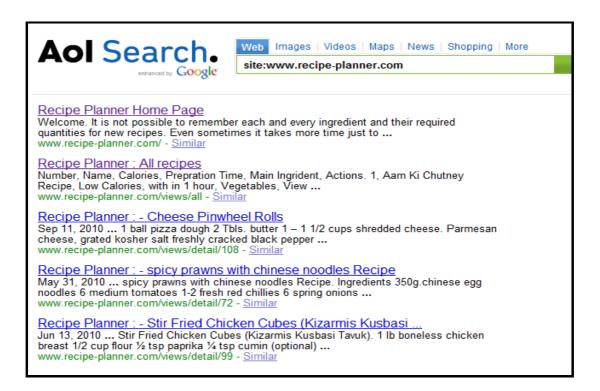


Figure 5.10 Posttest on AOL: Sitemap submission with friendly URLs.

The screenshot presented above is Posttest results i.e. after applying SEO technique (Friendly URL) of website2 on AOL search engine.

APPENDIX-B

SHAPIRO WILKS TESTS AND MANN-WHITNEY TESTS

The appendix-B is presented with Shapiro Wilks Tests and Mann-Whitney Tests

• Shapiro Wilks Test for indexing (recipe-planner.com)

Shapiro-Wilk W test for non-normality

Sample name: indexed Uncensored data = 5 Censored data = 0 Mean = 32

Standard deviation = 24.289916 Squares about mean = 2360

W = 0.686078P = 0.0068

Sample unlikely to be from a normal distribution

Shapiro-Wilk W test for non-normality

Sample name: not indexed

Uncensored data = 5 Censored data = 0 Mean = 99

Standard deviation = 24.289916 Squares about mean = 2360

W = 0.686078P = 0.0068

Sample unlikely to be from a normal distribution

• Shapiro Wilks Test for indexing (wasfabththesis.com)

Shapiro-Wilk W test for non-normality

Sample name: Start up
Uncensored data = 5
Censored data = 0
Mean = 0.2
Standard deviation = 0.447214
Squares about mean = 0.8

W = 0.552182P = 0.0001

Sample unlikely to be from a normal distribution

Shapiro-Wilk W test for non-normality

Sample name: After SEO

Uncensored data = 5 Censored data = 0 Mean = 9.4 Standard deviation = 2.50998 Squares about mean = 25.2

• Shapiro Wilks Test for Friendly URLs

Shapiro-Wilk W test for non-normality for Experimental Group

Sample name: Experimental Group

Uncensored data = 5 Censored data = 0 Mean = 7.2 Standard deviation = 5.674504 Squares about mean = 128.8

W = 0.734227P = 0.0211

Sample unlikely to be from a normal distribution

Shapiro-Wilk W test for non-normality for Control Group

Sample name: Control Group

Uncensored data = 5 Censored data = 0 Mean = 2.4 Standard deviation = 5.366563 Squares about mean = 115.2

W = 0.552182P = 0.0001

Sample unlikely to be from a normal distribution

• Mann-Whitney U test

Mann-Whitney U test for indexing recipe-planner.com

```
Observations (x) in indexed = 5 median = 23 rank sum = 16 Observations (y) in not indexed = 5 median = 108 U = 1 U' = 24 Exact probability: Lower side P = 0.0079 (H<sub>1</sub>: x tends to be less than y) Upper side P = 0.9921 (H<sub>1</sub>: x tends to be greater than y) Two sided P = 0.0159 (H<sub>1</sub>: x tends to be distributed differently to y) 96.8% confidence interval for difference between medians or means: K = 3 median difference = -85 CI = -95 to -30
```

Mann-Whitney U test for Wasfabththesis.com

```
Observations (x) in Startup = 5 median = 0 rank sum = 15 Observations (y) in After SEO = 5 median = 10 U = 0 U' = 25 Exact probability (adjusted for ties): Lower side P = 0.004 (H_1: x tends to be less than y) Upper side P = 0.996 (H_1: x tends to be greater than y) Two sided P = 0.0079 (H_1: x tends to be distributed differently to y) P = 0.88 confidence interval for difference between medians or means: P = 0.0079 (P = 0.00
```

Friendly URLs for recipe-planner.com

```
Observations (x) in Experimental Group = 5 median = 11 rank sum = 35.5 Observations (y) in Control Group = 5 median = 0 U = 20.5 \qquad U' = 4.5 Exact probability (adjusted for ties): Lower side P = 0.0397 (H<sub>1</sub>: x tends to be less than y) Upper side P = 0.9603 (H<sub>1</sub>: x tends to be greater than y) Two sided P = 0.0794 (H<sub>1</sub>: x tends to be distributed differently to y) 96.8\% \text{ confidence interval for difference between medians or means:} K = 3 \text{ median difference} = 1 CI = -1 to 12
```