Lappeenranta University of Technology

Faculty of Technology Management

Degree Program in Information Technology

Abrar Sohail

Search Engine Optimization Methods & Search Engine Indexing for CMS Applications

Examiners: Professor, Ph.D Kari Smolander
                      Professor, Ph.D Vladimir Dobrynin

Supervisors: Professor, Ph.D Kari Smolander

ABSTRACT

Lappeenranta University of Technology

Faculty of Technology Management

CBU Master Degree Program in Information & Communication Technology

ABRAR SOHAIL

Master's Thesis

75 pages, 14 figures, 9 tables,

Examiners: Professor, Ph.D Kari Smolander
                       Professor, Ph.D Vladimir Dobrynin

Search engine optimization & marketing is a set of processes widely used on websites to improve search engine rankings which generate quality web traffic and increase ROI. Content is the most important part of any website. CMS web development is now become very essential for most of organizations and online businesses to develop their online system and websites. Every online business using a CMS wants to get users (customers) to make profit and ROI. This thesis comprises a brief study of existing SEO methods, tools and techniques and how they can be implemented to optimize a content base website. In results, the study provides recommendations about how to use SEO methods; tools and techniques to optimize CMS based websites on major search engines. This study compares popular CMS systems like Drupal, WordPress and Joomla SEO features and how implementing SEO can be improved on these CMS systems. Having knowledge of search engine indexing and search engine working is essential for a successful SEO campaign. This work is a complete guideline for web developers or SEO experts who want to optimize a CMS based website on all major search engines.

ACKNOWLEDGEMENTS

I shall start in the name of Allah, the most merciful, who gave me the opportunity, resources and strength to complete this dissertation.

I would like to thank my supervisor Professor Kari Smolander for his kind support to accomplish my work. Special thanks to Professor Kari Smolander and Professor Vladimir Dobrynin on making this thesis better through discussions. I would like to thank my study coordinators Susanna Koponen and Riitta Salminen for their help regarding study and thesis matters.

I would like to present my love, respect and gratitude to my parents because of their prayers, wishes and kind support I am able to complete this task. I would also like to thank my friends and collogues who encouraged me during writing my thesis.

Finally, my special thanks go to my fiancé whose love and prayers give me energy to achieve all tough and impossible tasks.

Best Regards
Abrar Sohail

Table of Contents

TERMS AND ACRONYMS

API	Application Programming Interface
AJAX	Asynchronous JavaScript and XML
BP	Blog-Ping
CTR	Click through rate
CCK	Content Construction Kit
CMS	Content management system
CSS	Cascading Style Sheets
Cons	An argument or considerate in against/disagreement
DOC	Abbreviation for Document file
H1..H6	H use for Heading Tag
HTML	HyperText Markup Language
ECMS	Enterprise Content Management System
FFA	Free For All
GUI	Graphical User Interface
IP Address	Internet Protocol Address
IM	Instant Messaging
ID	Identification Number
IIS	Internet Information Services
I/O	Input/Output
IR	Information Retrieval
K.E.I	Keyword Effectiveness Index
K.O.I	Keyword Opportunity Index
MVC	Model View Controller
OPML	Outline Processor Markup Language
PPC	Pay Per Click
PR	PageRank
PHP	HyperText PreProcessor
PDF	Portable Document Format
ROI	Return on investment
RSS	Really Simple Syndication
Pros	An argument or consideration in favor
SEO	Search engine optimization
SEM	Search Engine Marketing
SERP	Search Engine Results Pages
SEF	Search Engine Friendly URLs
TLDs	Top Level Domains
TXT	Abbreviation use for Text file
URL	Uniform Resource Locator
WCMS	Web Content Management System

XML	Extensible Markup Language
WYSIWYG	What You See Is What You Get (A property of CMS)
404	Page Not Found error message
Adwords	SEM tool use for PPC campaigns
Analytics	SEO tool to analyze web traffic
ATOM	A XML feed language uses for Web feed
Dmoz	Open Directory Project of Google

1 INTRODUCTION

In this information and technology age online business, internet marketing and advertising have created a significant impact. Huge amount of revenues is generated by making and advertising good websites. However, there are billions of websites with vast range of categories and topics worldwide with distinct languages, areas and content types. To find information or website about a particular topic, a user uses search engine by writing search query in words, keywords, or phrases shape. In this era, most advanced search engines include Google, Yahoo, Bing MSN, AOL etc. There is huge development and research work has been done during the last decade to make perfect search engines for people in term of usability, reliability and finding right information according to the search requirements.

Further elaborating this trend search engines has become a vital need and daily tool for internet users and online advertising media. Today, internet market top Search engines are turning profit from advertising, entertainment, social media networks, daily use applications (e.g. maps) and online product sales and services. Internet advertising/ online advertising revenue in the United Estate totaled $26 billion dollars in year 2010 with increase of 15% over last year 2009. (Silverman, 2011) Internet traffic and business is growing day by day which indicates a massive growth in internet marketing and web development field.

Search engine optimization is the art of designing, developing, modifying and coding Web pages so that they can achieve high rankings in the search results and converting high user traffic. Search engines rapidly change their ranking algorithms with the passage of time but SEO basics remain same throughout. Search engine optimization can be performed by two kind of practices/methods which are White-hat SEO and Black-hat SEO. White-hat SEO called ethical or legal SEO in which activities carried out according to the guidelines, rules and policies of search engines and normally most of SEO practitioners follow it. Whereas Black-hat SEO refers to the illegal SEO practices against search engine rules and regulations which derived to get quick search ranking results.

There are billions of websites live on the internet. Most of them are based on managing, adding and modifying content. A content management system allows users to share, access control, contribute data and enhance communication between users in a collaborative environment. Joomla, Drupal, WordPress are the examples of today's most popular CMS applications. Most of CMS systems were not designed and developed according to all SEO standards. There are some specific problems with the content generated by every content management system.

Search engine indexing processes refers to the methods and techniques used by search engines to read and include your website content and data in their data repositories. Exploring the content of the Web pages for automatic indexing is the key importance for efficient ecommerce and other applications of the Web. It allows users, including customers and businesses to locate the best sources of their use. (Chung, 2001) It is important to keep index your all Web pages to get maximum ranking by search engines. If

your website is not indexed by search engine is like a huge shop without signboards and customers.

A search engine is program or software which fetches/retrieves data, files and information from a collective database or from the computer network. A Web search engine has three main parts. These are Web crawler, Indexer and searching. Web crawlers are programs that use the graph structure of the Web to move from page to page. In their infancy such programs were also called wanderers, robots, spiders, fish and worms, words that are quite evocative of Web imagery. (Gautam, 2004) Indexing is the process in which search engine store the crawled data with three processes i-e parsing, indexing document (storage) and sorting. The goal of searching is to provide quality search results efficiently. The goal uses a query evaluation process for searching. (Page, 2001)

This thesis report will comprise and evaluate detail research on search engine optimization methods for content management system applications and Web2.0. It also gives an overview of search engine indexing and how to improve search engine indexing for CMS applications.

1.1 Aims and Objectives

The main goal of the thesis is to research on latest search engine optimization techniques for content management based websites. Secondly, to analyze most popular content management systems performance regarding SEO point of view and how we can improve web development processes to make better CMS websites to make them search engine friendly and to generate high user traffic. Also, to achieve my goal I need to briefly study the search engine indexing processes and criterias for a website those are important for major search engines like Google, Yahoo and Bing MSN. It is important to study the results from researchers and practitioners, who are working in the area of search engine optimization-marketing, web 2.0 and CMS web development. To achieve my aim I have following objectives

1. To conduct a brief literature review on authentic material on given topic to get an overview of what has been done in area of SEO, CMS applications and search engine indexing.

2. To find and analyze new and existing Search engine optimization techniques and their affect on CMS applications and web 2.0.

3. To find the best SEO tools available those can be used to optimize websites on search engines.

4. Compare the most popular content management systems in this era, e-g Drupal, Joomla, WordPress with SEO perspective and analyze their performance in optimizing on top search engines.

5. To analyze and research on Google searching behavior and optimization standards for webmasters.

6.   To suggest and recommend solutions on existing layouts of different CMS based websites according to Web 2.0 and SEO standards.

1.2   Research Questions and Outcomes

A huge amount of literature in shape of books, articles, journals, authorized websites from pioneers of SEO and web development is available distinctly. The accessibility of most relevant, precise and fruitful information sources is a critical issue that always face when doing research on vast topics such as e-business, ecommerce and IT. This study in fact among such topics in which rapid kind of inventions and updates comes across on daily basis. This study comprises to describe and research all SEO theories at the moment and CMS web development with near future related technologies. The detailed study of literature available on search engines, SEO and CMS will help in conducting my research more effectively and efficiently. The study contains examples and essential solutions to run successful SEO campaigns and to build better, more optimized CMS applications in support to answer research questions. The main questions immersing from the research problem to achieve my aim and objectives which need to be answered are:

1.   What literature, latest SEO tools and techniques are available on SEO and CMS applications and how to integrate these SEO tools and techniques to optimize websites?

2.   How can development process of CMS based websites make better to improve search engine indexing?

3.   Comparison of the most popular available content management systems on basis of SEO performance and analyze which CMS is better and for which kind of application/system?

4.   How we can modify different CMS layouts of existing websites with respect to SEO and web 2.0 standards? It will include recommendations to get better SEO performance.

The study will briefly explain SEO methods for CMS applications and how to implement them for a successful SEO campaign and to generate high web traffic and revenues (ROI). This study will also comprise most reliable CMS application for the relative web application. The study will give an overview of search engine indexing techniques and how actually search engines rank websites and on which criteria are more critical for them.

1.3   Research Methods and Settings

The methods used for getting information and knowledge for this document are literature readings such like books, articles, research papers & journals, internet pages, technical white papers and experts views. The study is kind of review of existing literature in the field of search engine optimization/marketing, content management systems and search engine development and standards. The study mainly focus on web design and development particularly open source PHP covering area of content management systems,

latest SEO tools and techniques available and search engines. The study also includes criticism and analysis on basis of general defined facts on particular systems available and example websites picked for this thesis. The study include evaluation and analysis results with recommendations those can be used for any website without putting much effort to make it better for search engine.

This study is a descriptive, analytical, reviewing and doing postmortem on existing theories and knowledge in the field of web site design and development, internet marketing and search engine optimization processes. It also includes a review of search engine optimization and search indexing; furthermore it discovers strength, weaknesses, advantages and disadvantages of this field.

1.4 Structure of the thesis

The overview of the structure of thesis report is pictured in Figure 1.

Chapter1:Introduction & Overview of the topic

Chapter2: Literature review-: Search Engine optimization methods, tools and their working

Chapter3: Search Engine Indexing and Working Overview, Search engine indexing for CMS

Chapter4: Content management systems, Comparison of leading CMS systems on Basis of SEO

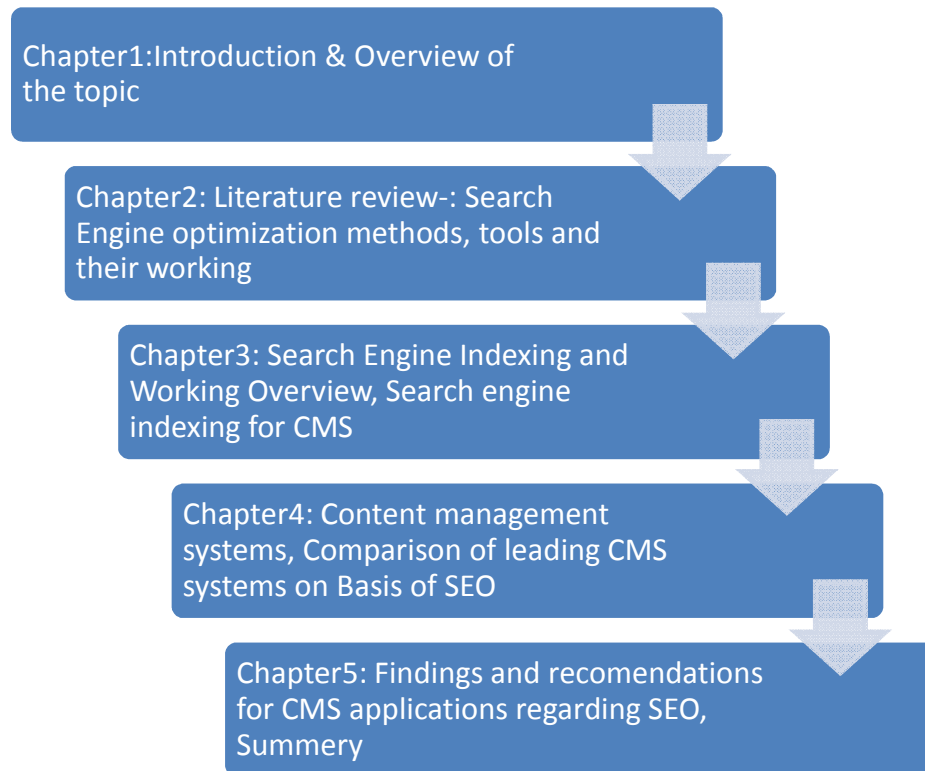Chapter5: Findings and recomendations for CMS applications regarding SEO, Summery

Figure 1 (Structure of the Thesis report)

2   SEARCH ENGINE OPTIMIZATION (SEO)

The importance of internet and research organization media as information source likes to rise up to conventional literature. For example, Google and other search engines rapidly change their search criteria and algorithms that's why search engine practitioners and researcher have to keep update themselves for latest theories and behavior of search engines. In this chapter I will present a detail overview of my thesis on topics such as search engine optimization methods, tools, techniques, content management basics and will compare their performance regarding search engine optimization and will also review of search engine working and indexing techniques. The primary search engine I focused is Google because now it becomes the king of search business and most used search engine now a days as well as other popular search engines like Yahoo, Bing and AltaVista.

Every one of us using internet opens Google or other search engine while finding information on internet. People who use the internet either looking for communication, information, entertainment or a product service like buying and selling. Search has become integrated into the fabric of our society and age. Likewise according to comSource, more than 12 billion searches being performed each month as of January 2009, approximately 400 million web searches performed every day. (Enge, 2009) Search engine optimization is a vast term which covers a huge area on internet. Organic search are listings on the search engine which are results of search queries when a user writes keywords in search engine while searching. And those are not includes adverts sponsored links (pay-per-click ads) (James, 2011)**.** A website can only be optimized and visible when it is well ranked and placed on first page in search engine results. *Search engine optimization* is a process and set of theories, tools and techniques apply to get website ranked and placed on first page on organic search. It means if any user is looking for a website relevant to your keyword/search term so that your website should come on first page (most likely in first three positions) of search listings which enable the maximum chance to have a click on your website.

The perfect way for your optimized website is that it should be displayed on first page with top positions of search results/listings. It has been researched and surveyed that most of the people click on first five search listings results and don't to look beyond the third page if they go for maximum search results. That's why those website who occupied first five positions have maximum traffic and they generate more profit. There are several ways and practices to achieve search engine optimization and most of the people apply them in different ways, methods and sequences, but the general practices and guidelines of SEO are same for all kind of website campaigns. For example, a major search engine like Google has given the guidelines in their help pages to help webmasters and SEO practitioners. If the general guidelines and SEO practices are not followed then these can cause failure of SEO campaigns, spam or illegal for search engines.

2.1 WHY Need of SEO?

Search engine optimization and marketing is now become significant need of every online business, product and services. Many online businesses are not successful because they are not optimized well on the search engines. Using paid advertisement such as Google Adwords still need SEO skills to write optimized ads. Search engine optimizer/practitioner emphasis not only website structure, content, design and code but also take massive attention on building external links on regular basis. The goal of search engines is to provide quality content, fast search results and advanced easy advertisement opportunities to people searching on the internet. The more times search results lead to the desired content, the more likely a user use that search engine again. Generally, a well optimized website must be user-friendly, well-structured and displayed on first page on its main keywords. With these entire characteristics website get more traffic which ultimate increase the ROI.

Search engine optimization depends a lot upon the website development factors because if the website is not developed well enough according to search engine guidelines cannot be optimized well for search engines. Today the invention of content management system (CMS) took big part of web development as now almost every online business and corporate website based on content management system. Web2.0 is a second generation standard platform for designing and developing. Web applications make easy communication, interoperability, share information, user-centric collaboration and integration on the World Wide Web. Most of social networking and highly web traffic websites are based on Web2.0.

Search engines are designed for people to seek information and search on internet. Search engine provides a wide range of applications and tools for various purposes like advertisement, search engine optimization and informational. Every crawling search engine has three parts. These are crawler, indexer and a search interface. It is important to study how a crawling search engine works and what are the main indexing techniques used by major search engines like Google, Yahoo and Bing. There are various techniques and tools to optimize a website. It is essential to find which SEO techniques and tools are critical for every SEO campaign and which standards and guidelines should be followed while developing a CMS based website. There is distinct range of CMS tools available on internet open source market. From these, Drupal, Joomla and WordPress are leading shareholders.

2.2 SEO Strategy

It is better to make a brief SEO strategy, work accordingly and in sequence, timely and goal driven. It is also important to keep update yourself about the latest ranking strategy and updates of search engines. Effecting SEO strategy requires iterative testing in order to refine it. Like advertising, SEO needs long term dedication in order to see best desired results. There are four general steps to a refined SEO effort; (Potts, 2007)

1. Strategize

2. Execute
3. Analyze
4. Optimize

In first step you make a specific SEO strategy which is the part of SEO planning process. In second step you execute the essential steps which are predefined in strategy. Then you analyze the SEO process and finally you get in optimization process and see the results. SEO strategy is not unique to all types of websites, for different kind of websites the SEO strategy will be different and also the time frame of the SEO campaign to show the desired results. SEO strategy is further divided into its phases with respect to SEO components. All of these strategies can be rolled up to single artifact which is SEO plan. These can be content strategy, link building Strategy, search engine marketing strategy, social media strategy, search engine targeting strategy, technical strategy etc. (Jerkovic J. , 2009).

2.3   SEO Goal and Target Audience

After making a brief SEO strategy it is time to define goal and your target audience. The main goal of SEO campaign is to get website high up in the search results on its target keywords. In this process you define your target audience, what clients want to sell or promote, which sort of visitors will be target, which area or country and time frame to complete it. It is very important to measure your goal and efforts and what outcomes are in resulting of your SEO efforts. While web traffic measurement relates to the optimization effectiveness measures, goal means more to the business effects of SEO. Goal measurement is essential because it is completely possible to increase traffic while hurting your business. (Tonkin, 2010) It depends upon the product or services you are offering. For example, a restaurant in Rome, Italy business has first goal to get ranked locally search engine and to get customers from Rome, Italy area. The basic ingredients of search engine optimization are:

- Quality content
- Relevant keywords
- Strong metalanguage properties,
- Internal links throughout the site,
- External links (backlinks) to the site. (Gifford, 2010)

2.4   Domain Name & Hosting

To start an online business the first thing you do is to register a domain name. After knowing you target area and target keywords try to register the domain name which includes your primary keyword. For example, if your business sells laptops then the domain name with keywords laptop (e-g buylaptops.com) will be a good domain name. Choosing webhost is also an important issue because while editing your website for SEO requirements, webhost should offer htaccess[1] file modification, SEO add-ons (SEO plugins integration) and facility to buy a dedicated IP for your domain name.

---

[1] A **.htaccess** (*hypertext access*) file is a directory-level configuration file supported by several web servers, that allows for decentralized management of web server configuration.

Following are the key elements to look out when registering a domain name: (Danny, 2011)

Avoid hyphens: In domain names, hyphens detract from credibility and proceed as a spam indicator.

Avoid generic, uncommon top-level domains (TLDs): Like hyphens, TLDs such as .info, .cc, .ws and .name are spam indicators.

Longer Domain Name: Avoid domain name longer than 15 characters.

Beware of permutations: The owners of ExpertsEchange.com built a sizable brand they recognize their domain name could be misconstrued as Expertsexchange.com (Danny, 2011) it means that expertsexchange.com can also be misunderstood by "Expert Sex Change" which can refer to adult content or some other industry which can mislead user's interaction. That's why Experts exchange a leading IT helpline company has domain name "experts-exchange.com"

The age of the domain can play a vital role in assigning of Google PageRank. Young domains are normally ranked low; consequently older domains are always ranked higher. However, if you buy a recently expired domain name that has a high PageRank because of its age and accumulated reputation, the PageRank does not transfer to the new owner. Search engines can generally see when domain names change ownership, and will reset their rankings in such situations. (Walter, 2008)

2.5 Keyword Research/Discovery

The most common type of internet search is keyword search. Keyword search refers to a type of search in which keywords are typed in a search box to locate information on the internet. (Morley, 2009). To perform a keyword search a user enters keywords or phrases in search engines. For example a person looking for listings of web host might type 'web host' in search box. Before selecting a domain name for your business it is recommended to perform a keyword research process. Before starting the keyword research it is necessary to know the main subject (keyword), area and competitors of your website. Collect the words that best describe your website and its contents. These words will be your primary keywords, from these you can make combination of keywords and phrases to use in SEO campaign. For example if the topic of the website is web hosting. Then the keywords like 'cheap web hosting', 'best web hosting' and so on can be your targeted keywords. In keyword research process you need to go thorough following steps

- Understand your target market and make a note of that market which you are targeting.
- Create a general list of keywords and keyword phrases.
- Extract keywords from competitor's websites and target market which is important in your point of view.

- Use appropriate keyword research tools available on internet such as Google, Keyword Discovery, SEOBOOK and WordTracker having best KEIs and KOIs.
- Assess the popularity of your keyword choices. How many site matches show for your keyword searches? From the top 30 search listings how many have optimized their websites? So how you can do better? (Rice, 2009)
- For each page of your website use unique keywords according to the page content.
- Keyword research is very important and initial phase in any SEO campaign.

2.5.1 K.E.I (Keyword Effectiveness Index)

Keyword effectiveness index (K.E.I) firstly was created by Sumantra Roy. (David, 2008) It is the ratio of search count (number of search performed by users) to the total number of search results while search that keyword. In simple way KEI compares the number of searches and the number of web page results respectively. The search volume can be known by using keywords tools like WordTracker, Google Keyword Tool, SEOBOOK and many on the internet. These tools give estimated figure of searches done by users in a particular period on search engines. (See Figure 2)

Try The Coolest Keyword Research Tool Today!

Enter your keyword. This tool is powered by Wordtracker.

web hosting   ☐ Show Google Insights

Submit

WordTracker	WordTracker count	Google daily est	Yahoo! + Bing daily est	Overall daily est	Yahoo! Suggest	G Trends G Insights	G Suggest	G SB-KW
web hosting	8376	10,470	4,314 YI B	14,772	YI Sug 2 Clues	G Trends G Ins	G Sugg	G SB-KW
hospedaje web hosting	2404	3,005	1,238 YI B	4,240	YI Sug 2 Clues	G Trends G Ins	G Sugg	G SB-KW
free web hosting	1957	2,446	1,008 YI B	3,453	YI Sug 2 Clues	G Trends G Ins	G Sugg	G SB-KW

Figure 2 (SEOBOOK Keyword Research Tool showing daily estimated search on "web hosting" keyword (www.seobook.com keyword research tool)

There are various ways to calculate KEI. SEO practitioners have introduced different kind of KEI formulas. The simplest way to calculate KEI is number of searches divided by number of search results. Suppose the number of search count for a keyword is 2000 per month and Google displays 150,000 results for that keyword. Then competitiveness for that keyword is 2000*2000/ 150000 which results 26.6. Hence the general formula to calculate KEI is

$$KEI = (Monthly\ Searches)^2 / Raw\ Competition\ (Keil, 2011)$$

Where "Monthly searches" refers to the estimated number of search queries done by people in one month. For example, according to figure 2, estimated monthly search for keyword "web hosting" is around 14,472. Raw competition is the total number of SERPs for a respective search query. For example, total SERPs for "web hosting" keyword on Google search engine is 455 millions websites.

However, according to SEO expert David Viney the results of the KEI formula seems to be diluted when using the total Google page results count by taking any web pages into account despite of purpose and relevance of the listed websites. (Keil, 2011)

2.5.2   K.O.I (Keyword Opportunity Index)

Keyword opportunity Index (KOI) refers to find out which keywords are likely to succeed and its measure the attractiveness of the keyword on directly competing site. To get the directly competing websites the query for Google search engine is "allinanchor: keyword" Where anchor represents the websites with anchor text links contain your targeted keyword. For example for keyword "web hosting" direct competition websites are 193,000,000. The formula of KOI is

$$KOI = (\text{Monthly Searches})^2 / \text{Directly Competing (Keil, 2011)}$$

The reason of Google search command "allinanchor: search term" is to get web pages in order of the highest number of inbound links that contain the queried keyword in their anchor text links. The anchor text inbound links are important indicators for the Google's algorithms and how relevant the site is to the queried search term. (Keil, 2011) Example of calculating KEI and KOI is shown in table 1.

Keywords	Monthly Searches	Raw Competition	Directly Competing	KEI	KOI
Business Cards	214,249	245,000000	365,000	187.53	125,878.07
Business Card Printing	19,524	42,100,000	36,600	9.05	10,414.93
Business printing	6,345	127,000,000	120,000	0.32	335.49
Online business card printing	1,265	516,000,000	13,300	0.00	120.32

Table 1(Example of Calculating KEI and KOI **(David, 2008)**

2.5.3   Keyword Density

Keyword density is the metric used in keyword analysis that defines the ratio of the number of occurrences of a particular keyword or phrase to the total number of all words on a given web page. (Hayman, 2007) If a keyword or keyword phrase appears many times in a single page (i-e has high keyword density) then search engine will consider it spam. The ideal way to use keyword density is between 4% to 6%, sometime is said to be as high

as 10% but it is not recommended. For example, if a web pages has 500 words and if your target keyword or phrase 50 times in that page, then the keyword density of the particular keyword is 50/500*100 which is 10%.

HTML	Keyword	Total	Density (in percentage %)
Title	3	10	3
Meta_Description	4	18	2.22
Meta_Keywords	8	26	30.76
Visible_Text	29	397	7.3
Alt_Tags	1	31	3.1
Comment_Tags	4	42	9.2
Domain_Name	0	1	0
Image_tags	0	81	0
Linked_Text	78	78	100
Option Tags	0	0	0
Reference_Tags	0	134	0
Total	127	818	15.52

Table 2 (Keyword density sample report source: keyworddensity.com)

2.6 Competitor Analysis and why it is important?

Before starting search engine optimization campaign it is very essential to know about SEO tactics using by your competitive websites to attain ranking. In every business trend competitor analysis is a key activity. Knowing the behavior and tracking the competitor will help to implement SEO techniques more efficiently. It is need to assured that your site will be looked at, crawled and scrutinized by your competitors or their SEO practitioners. (Jerkovic J. , 2009) The first thing you need to evaluate is which are your competitors, and on which keywords they are competing for. After conducting the keyword analysis extract most relevant and targeted keywords. Using these keywords in search engines and note down the top websites come in search results. Also you can get your competitors by content by writing this query" related:www.yourwebsite.com" in Google search engine.

The key factors in competitor analysis are

- *Competitor website indexing:* Analyze how much pages of your competitor websites are indexed by Google and other search engines?
- *Link popularity and Page Rank:* Back links and page rank of the competitor website.
- *Targeted Keywords:* what are their main keywords and main page title and Meta tags?
- *Estimated Web Traffic:* Analyze the web traffic they are getting. To get traffic data there are several tools are available on internet. The most common are Alexa.com

and doubleclick.com. Following table3 is showing a sample of competitor analysis report.

URL	Last Updated	PR	Alexa Rank	Google Links	Yahoo! Links	MSN Links	Outgoing Links
http://www.jlentertainment.com	October 28th,1999.	4	746,128	128	20,260	1	11
http://www.delafont.com	25th June, 1997	5	199,947	212	2,291	2	35
http://www.bookingentertainment.com	November 17, 1999	4	253,833	152	6,122	0	68
http://www.universalattractions.com	February 01, 1997	3	2,473,656	5	114	5	39
http://www.ejams.com/agency	March 17th, 2008	3	224,038	47	13,013	2	133
http://www. wandmtalent.com	March 10th,2009	3	3,292,542	2	7	0	25

Table 3 (Competitor analysis sample report)

2.7 Search Engine Crawling & Indexing

Crawling is performed by robots (bots) which are also known as search engine spiders. The main function of the crawl is to classify relevant pages for indexing and evaluate whether they have changed. Search engine crawlers and spiders access the web pages and retrieve a reference URL of the page for later analysis and indexing purposes. (Dave, 2009) The data collected by spiders and crawler is used to display search results and search engine ranking by search engines. Search engines crawl the websites on regular basis depends upon the website progress, quality and ranking. Websites who update frequently their pages with updated content on daily basis are crawled and indexed more than an ordinary postdate content. Web crawler used to store a copy of those web pages those are crawled. It can be seen by writing cache: with the following URL address need to be checked in Google search box i-e write this query in Google search box

cache:www.yoururl.com

The screenshot of this process is shown in following figure 3.

Figure 3 (Google Crawling process)

 A web page when cached by search engine (e.g. Google) is said to be indexed. Indexer is particularly designed and optimized for indexing files. Using the index built by the indexer, the search engine can access almost directly to sections of the database which contains the information a user is looking for. (Stanek, 2001) Search engine ranking depends a lot upon the website indexing. The more of website's web pages include (indexed) by search engine then it will have better search engine ranking. Search engine practitioner's one of primary purpose is the website indexing so that every desired web page should be index. The detail of the search engine indexing is discussed in search engine indexing section.

2.8   PageRank and its Importance

PageRank is a link analysis algorithm used by Google search engine originally formulated by Larry Page and Sergey Brin. The PageRank values are pre-calculated and stored for all pages known to the IR system. This means every page in the web has a PageRank score that is entirely independent of query terms. A search that returns PageRank scores is reporting the significance hierarchy of pages containing the query terms. (Thorson, 2004) It is a value in number out of ten (10) which Google toolbar gives on basis of numerous factors and can be seen using Google toolbar or SEO tools.

PageRank is an exceptional way to prioritize the results of web keyword searches in order. For most admired subjects, a simple text matching search that is restricted to web page titles carry out commendably when PageRank prioritizes the results. The major factor PageRank depends upon is the back links to a website. PageRank widens this idea by not

21

counting links from all pages equally, and by normalizing by the number of links on a page. PageRank is defined as follows: (Page, 2001)

Let assume that page A has pages T1...Tn which point to it (i.e., are citations/back links). The parameter d is a damping factor which can be ranged between 0 and 1. Generally, assumed value of d is set to 0.85. Also C(A) is defined as the number of links going out of page A. The PageRank of a page A is given as follows: (Page, 2001)

PR(A) = (1-d) + d (PR (T1)/C(T1) + ... + PR(Tn)/C(Tn))

It is important to consider that PageRanks form a probability distribution over web pages, so the sum of all web pages' PageRanks will be one. (Page, 2001)

The Google toolbar is not very correct in telling you the PageRank of a site, but it's the only thing right now that can really give you any idea. There are two limitations to the Google toolbar:

1. If you enter a page, which is not in its index, but where there is a page that is very close to it in Google's index, then it will give a guesstimate of the PageRank. This guesstimate is worthless for our purposes because it isn't featured in any of the PageRank calculations. The only way to tell if the toolbar is using a guesstimate is to type the URL into the Google search box. (Chris, 2001)

2. The true PageRank can be seen by installing Google toolbar (i-e. from *toolbar.google.com*) in the web browser. PageRank is linear, so Google use a non-linear graph to show it. To get from a PageRank of 2 to a PageRank of 3 needs less of an increase than to move from a PageRank of 3 to a PageRank of 4. The actual figures of PageRank is kept secret however following figure 7 is the estimated actual figures corresponding to the Google toolbar PageRank (Chris, 2001)

If the Actual PageRank is between	The Toolbar Shows
0.00000001 and 5	1
6 and 25	2
25 and 125	3
126 and 625	4
625 and 3125	5
3126 and 15625	6
15626 and 78125	7
78126 and 390625	8
390626 and 1953125	9
1953125 and infinity	10

Table 4 (Actual PageRank corresponding to Google toolbar PageRank (Chris, 2001)

PageRank factor has high importance in buying/selling links. Websites with high PageRank demand good money to give inbound links, for example Apple.com has PR 9 so having back link from apple.com values a lot for search engine. However, it is not

necessary that if a website has high PageRank can have better search ranking and high web traffic comparing a website having a low PageRank.

Google never disclosed the actual PageRank algorithm calculation which determine by Google toolbar however PageRank depends upon the following factors

- ✓ The amount of back links to the website, quality and content relevancy.
- ✓ The PageRank of the websites is linking to your website.
- ✓ The amount of outbound links on the page is linking to your website. For example, a web page has 10 outbound links and has PageRank 4. The Google divide the PageRank to all ten web pages according to their relevancy and authority factors.
- ✓ The age of the domain.
- ✓ The amount of click through rate on a website in query search.

3 ON-PAGE OPTIMIZATION

On-Page optimization also named on-site optimization is about making changes and what you do on website in its design and development process to improve and experiment on SERPs. It also includes in critical planning steps like understanding your niche, keyword research and SEO web strategy. (Fleischner, 2011) On-page optimization process includes those components directly related to your website

- *Code optimization:* It refers about modifying and adding page code, title & metadata, alt, heading and miscellaneous tags.
- *Content:* It is about how website content will be displayed? It includes keyword rich content, tweaking, and keyword density factors.
- *Directory/Link structure:* It includes directory/URL structure, breadcrumb trail, URL re-writing factors.

3.1 Code Optimization
3.1.1 Page Title

For search engine rankings, title tags are the most critical element for the search engine relevance. The title tag is in the <head> section of the HTML document, and this the only part of "meta" information about a page that influences the most relevancy and ranking. (Eric, 2009) It represents the topic and main keywords of the particular web page. It is the combination or string of the words/text defines by the tag <title> in HTML document. Following is the syntax of Title tag in HTML

<center><title> Your web site Title </title></center>

The title of the page is visible both in title bar in web browser and in headline of the search results. It is recommended that website title should be in range of 50-80 characters maximum in length including spaces. Google shows in average about 55-70 characters in organic search results. In search engine marketing web page or landing page title plays a very important role to boost the CTR (click through rate) of the website or a particular advert. CTR of the website is computed by dividing the number of clicks an ad/page got by the total impression bought. (Bidgoli, 2010) For example, if an ad shown by 10000 times and gets 100 clicks then CTR is 1% (100/10000).

3.1.2 Description Tag

Web page Meta description tag contains brief concise information about the page content. Normally it is displayed after title in search engine results page (SERP). The Meta description tag is important because you can utilize it as a method to convey your marketing message and entire search engine visitors to click on your listing versus clicking

your competition. (Kristopher, 2010) The Meta description tag is in the <head> section after title tag. The syntax is as follows:

<Meta NAME="decription" CONTENT"your website description.">

In Meta description tag you can use your targeted keywords which is important to increase CTR and for better SEO. Generally, Google shows about 155 characters in SERPs. Meta description tag is ideal for search engine when it is 150-200 characters long in length.

3.1.3   Meta Keywords Tag

The meta keywords tag offers you to present additional keywords or text for crawler based search engines to index along with your body content copy. When writing Meta keywords tag, always write most essential keywords first, as they hold the most relevance for search engines. (Michie, 2007) Latest research says that Meta keywords tag has not much importance for major search engines such like Google, but still it is affective for many other commercial and small search engines like AltaVista, AOL etc. It is ideal to you use 4-5 keywords in the meta keywords tag. The syntax is as follows

<Meta NAME="keywords" CONTENT="list of keywords separated by commas (,). ">

3.1.4   Meta Robots Tag

The robots tags specifically use to control search engine indexing process for a particular page. It specify whether a web page of website should or should not be indexed by search engine and also controls the links (text links) in the website whether they should index by search engine spiders and crawlers. Normally, in robot tag there are two parameters and sytax is as follows

```
<META NAME="ROBOTS" CONTENT="ALL | NONE | NOINDEX | NOFOLLOW">
      default = empty = "ALL"
      "NONE" = "NOINDEX, NOFOLLOW"
```
 Where

     Index: It allows GoogleBot to index that page.

     noindex: It tells GoogleBot not to index that page.

follow: This allows GoogleBot to follow the text anchor links from the landing pages and other pages (external links to other websites)

nofollow: This allows the search engine spiders not to follow links from that page for indexing.

Noodp: Avoid the search from using the page's description. (Odom, 2011)

Terms	GoogleBot	Slurp	MSNBot	Teoma
NoIndex	YES	YES	YES	YES
NoFollow	YES	YES	YES	YES
NoArchive	YES	YES	YES	YES
NoSnippet	YES	NO	NO	NO
NoODP	YES	YES	YES	NO
NoYDIR	NO	YES	NO	NO
NoImageIndex	YES	NO	NO	NO
NoTranslate	YES	NO	NO	NO
Unavailable_After	YES	NO	NO	NO

Table 5 (A quick reference of Meta robots tag's Usage (Seo, 2011)

Where

Googlebot: Google's web crawling bot (sometimes also called a "spider")
Slurp: Slurp is a web crawler from Yahoo search engine.
MSNbot: MSNbot is web crawler of Bing MSN search engine.
Teoma: Refers to crawler of Teoma search engine, which is powered by Ask.com (A popular search engine).

3.1.5   Heading Tags

Heading tags are very important as these are headlines of the topic of the website. Heading tags are given more worth by search engines than a regular body copywriting. Using your targeted keywords is very important in heading and subheading tags. Normally heading tags are written from H1 through to H6. Heading tags tells the search engine robots about the topic of the website. The H1 tag has the most importance and H6, the least and vice versa. Keeping code to H1 through H3 heading level is an ideal approach. The syntax of heading tag is as follows:

<h1>Heading Title</h1>

<h2>First subheading</h2> and so on.

3.1.5.1 Keyword tweaking

It is term used for making a specific content bold or italic. While writing the content of the website, some webmasters make important keywords bold or italic to improve search engine ranking. Keeping bold is also useful to draw attention for the reader.

3.1.6 Breadcrumb Trails

While designing the structure of the website breadcrumb trials plays an important role because it makes website easy to understand and search engines to follow the website directory structure. A breadcrumb trail is a text based navigation which shows the hierarchy of the website categories and URLs. It provides a major advantage for SEO if the links are text anchor links and present by efficient keywords. A general structure of writing breadcrumb trail is as follows

Home Page » Section/Group name » Category name » Page name – Page description

For example a breadcrumb trails for a web hosting site can be

Home > web hosting > cheap webhosts > cheap web hosting list

Breadcrumb trails is very important in SEO due to following three reasons (David, 2008)

- ➢ Breadcrumb trails support the navigational structure of the website so that PageRank is distributed evenly down through your pages.

- ➢ Breadcrumb trails are normally the first text that appears in the body of the web page and thus using your best keywords according to URLs will be affective in SEO process.

- ➢ Back links to web pages that integrate keyword-rich anchor text are always of value, even when those links come from other source in your own website. The breadcrumb trails offers you a legitimate opportunity to take in more of these. (David, 2008)

3.1.7 ALT Tags

Search engines while indexing do not read the images and graphics such like flash used in the websites. It is very essential to utilize your web site ranking in particularly in image search results. Bing (MSN) search engine shows images in its regular search results and image search and relies significantly on the alt tags in its algorithms. Google also gives string importance to the alt tags. (Evan Bailyn, 2011) It is better to use keywords and phrases in alt tags relevant to image and page content so it will appear in image and regular search. The syntax of Alt tag is as follows:

For example,

3.1.8 Using Sitemaps

Sitemaps basically a list of all URLs in a website which includes web pages URLs and other files e.g PDF, TXT, or DOC files. There are two kind of sitemaps normally use in a website. These are HTML or any other scripting language written sitemap (e.g sitemap.php) and XML sitemap. XML sitemap is used to submit in Google tool called Google webmaster[2] so that Google will index all the web pages listed in that XML sitemap.

Sitemaps offer search engines to more intelligently crawl your website and also allow you to control how search engine prioritize your web pages. Furthermore, they allow you to specify the frequency of change your pages undergo; offering search engine robots to better manage revisits. (Blankson, 2008)

The sitemap must include (organization, 2008)

- A <urlset> tag and end with a closing </urlset> tag.
- Must indicate the namespace (protocol standard) within the <urlset> tag.
- A <url> entry for each URL, as a parent XML tag.
- A <loc> child entry for each <url> parent tag.

The syntax of the writing xml sitemap is as follows: (organization, 2008)

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>http://www.example.com/</loc>
    <lastmod>2005-01-01</lastmod>
    <changefreq>monthly</changefreq>
    <priority>0.8</priority>
  </url>
</urlset>
```

3.1.9 Directory/URL Structure

One of the most discussed topics in search industry is importance and usage of Web page's URL structure for search engine ranking. The best approach is to develop Web pages URLs simple and static. If the URL contains relative keyword to its page content and it can easily be crawled, then it will be better to get top search engine ranking.

A site's URL structure should be as simple as possible. Thinking about organizing your content so that URLs are developed logically and in a way that is most clear to humans (when possible, readable words rather than long ID numbers). For example, if you are

---

[2] Google Webmaster Tools provides you with detailed reports about your pages' visibility on Google. It is used to verify website according to Google guidelines and to monitor the SEO performance.

searching for information about cars, a URL like http://en.wikipedia.org/wiki/Cars will help you make a decision whether to click that link. A URL like http://www.example.com/index.php?id_sezione=260&sid=4a6ebc123f22ada7gf49f521f1 is much less attractive to users. Secondly use punctuation in URLs. (Google, 2011)

For example, the URL http://www.example.com/ford-cars.html is much more useful than http://www.example.com/fordcars.html. Using (-) is better than underscores (_) as well as Google consider fordcars a single word. (Google, 2011)

The key factors should consider while developing a website are as follow: (Danny, 2011)

- ✓ Homepage links to every category of pages on the website
- ✓ The presence of links appear in category pages to all relevant subcategories not exceeding the standard limit of inbound links.
- ✓ The presence of links on subcategory pages to all relevant content pages.
- ✓ The ability of the URL structure to match category hierarchy and supplement relevancy. (Danny, 2011)
- ✓ Avoid dynamic URLs which contain symbols (e.g session IDs,? =, %, &, * and characters).
- ✓ Make website page URL simple, static and according to W3C guidelines.
- ✓ It is recommended that directory structure of the website should not be more than third level which means URL of the web page should not exceed more than three slashes (/) otherwise it will be difficult for search engine to read the URL.
- ✓ It is ideal to give one click access approach to the home page in deeper website pages. Home page of the website has maximum search worth for search engine. Keeping this factor will boost the index process of website.

3.1.9.1 Outbound links

Outbound links are the links to other sites present on your website. Search engines give importance to outbound links depending upon the quality, PageRank and content relevancy factors. Ensure that all outbound links in your site are meaningful and not spam. They contain relevant keywords and take the visitor to your related content page. This applies to search engine that all links are relevant to your site content and also to visitor to find the appropriate information. (Small, 2005)

4 OFF-PAGE OPTIMIZATION:

Off-page optimization also called off-site optimization is refers to the factors that have an outcome on your website ranking in natural search results directly related with outside circumstances (consists search ranking factors not located on your website). It is divided into two major components. These are website's history and Links back (outbound links) to the website. The most important factors of off-page optimization include the following: (II, 2008)

- ✓ The amount of websites which linking to your website i-e outbound links.
- ✓ The PageRank and link popularity of the websites linking to your website
- ✓ Typical and content relevance of website linking to your website
- ✓ The anchor text used in the back links (outbound links)
- ✓ The quality of the Web pages linking to your website which include total number of links on that Web page, page title and content.
- ✓ The IP address of the website linking to your Web page.
- ✓ Directory and search engine submissions.
- ✓ Having a proper robot.txt file.
- ✓ Google Sitemap formation and submission via Google webmaster.
- ✓ RSS syndications and many other factors. (II, 2008)

4.1 Link Popularity

Link popularity refers to the total number of web sites that link to you website, in other ways link popularity is the combination of all kind of back links (e.g. from websites, forums, blogs, RSS etc) to your website. It also includes the popularity of the web sites as well as content relevancy of those linking to you.

4.1.1 Link building

Link building is a process to create inbound links (website linking to your website). These links can be generated by being listed in search engine directories e.g. Dmoz, yahoo directory, relevant topic websites, search engines, newsletters, press releases, adverts, blogs, forum and social media networking sites.

Link building is vital in search engine optimization process and a key element in off-page optimization. Conducting a link building campaign will help to improve web site's link popularity, increase the website traffic and eventually improve the search engine ranking. Links are not equal; every website on the internet has some distinct worth and intrinsic value depending upon its search ranking factors. Sites that are linked to high-authority sites

turn into higher in the authority chain themselves. (Rice, 2009) There are various methods to build links which are discussed as follows

4.1.1.1 Reciprocal Linking

A reciprocal link is a mutual way of link building in which two websites swap their links to each other. For example, Website www.a.com has an inbound link to www.b.com in exchange of the link of www.b.com. Reciprocal links has the least worth in terms of search engine rankings however, it is effective to get rank and boost web traffic.
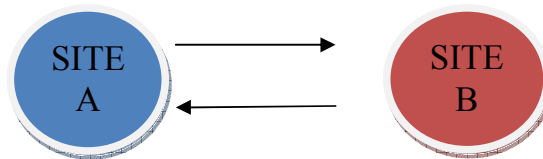

Figure 4 (Reciprocal Linking)

4.1.1.2 One Way Linking

One way linking is the most effective way and most valued. One way link is when another website is linking to your website and you have not linked back. Normally, websites do not give link easily. Usually the process of one way linking is about buying links. Website links from high PageRank website are more expensive.
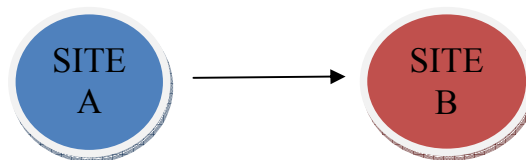
Figure 5 (One-Way linking)

4.1.1.3 Three way or Triangular linking

In three ways linking, site A gives links to site B, and site C links back to site A. For example, site a.com wants a 3 way links from site b.com then site a.com will add link to b.com on its partner or directory site which is c.com and will ask a.com back link on site b.com. In this way both site a.com and b.com get a one way link. This way of linking also called triangular linking.
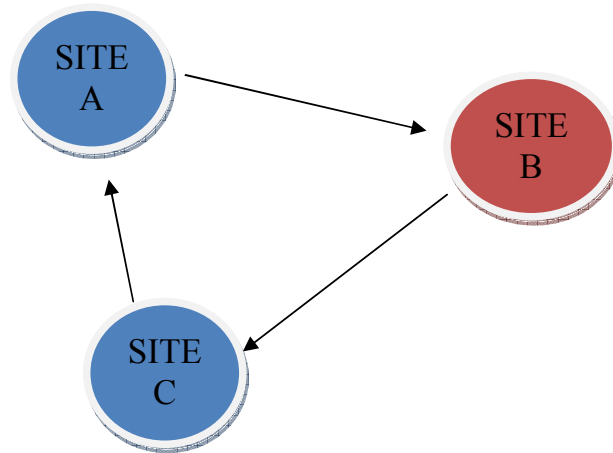
Figure 6 (Three way or Triangular linking)

4.1.1.4 Link Baits

Link baiting illustrate the concept of creating interesting content and text anchor links back with the intent of creating maximum internet buzz to get as many people to link to your sites as possible. It exists in many forms. Having great content is most basic and popular link baiting technique. (Jerkovic J. , 2009) For example, if a site owner buy link bait for his hosting site from a blog. The content of link bait with text anchor links can be as follows:

"Find cheap web hosting from world top web host at affordable price. "

Generally, blogs are popular in link baiting business. Link bait is a very effective way to get search engine rank quickly.

4.1.1.5 Search Engine Submissions

After making ready your website, it must be submit to all major search engines as well web directories e.g. Dmoz, Yahoo, Business etc. Most of the search engine submission pages can be found by writing query in Google search engine "add url". It is important to understand that search engine optimization services are application-based or online. There are many paid services available as well. It is important to notice online services that promise to submit your website to "thousands of search engines" because it can be spam. It means that your URL will be submitted to "link farms" and "Free For All (FFA)" pages as well legitimate search engines. (Kernek, 2005)

4.1.1.6 Improve Click Property

Click popularity is a very important factor for website rank in search engines. The popularity of the page regularly clicked will be high. When a visitor clicks a website from SERPs, search engine will offer some certain value. But do not try to replicate clicking on your own website, because the clicks from a single IP will be counted only once. (Saini, 2011)

4.1.2 Google SEO Tools

Google search engine hold the 65% of the world search market. There are some effective search engine tools introduced by Google. Every search engine practitioner must works on Google webmaster, Google analytics, Google trend, Google keyword tool in the search engine optimization process.

4.1.2.1 Google Webmaster Tool

All major search engines like Google provide free tools for webmasters and SEO practitioners. Google webmaster tool is basically to manage website on-site and off-site SEO performance.

Google webmaster can help your site performance better in search results as well as gives the facility to update your website indexing and crawling process concerning site's URLs. The major functionalities offer by Google webmaster tool are

- Uploading Sitemap in XML format and verify website by Google with a unique ID.
- Submissions of robot.txt file which allows blocking URLs which you do not want to index by Google.
- Removing URLs which are already crawled by Google.
- Indicate the chosen/preferred domain.
- Identify problems with title and meta tags.
- A detailed report containing the data of impressions, keywords and respective CTR.
- Data of all indexed web pages by Google with date.
- Back links (inbound links) to your website with crawling date by Google.
- Obtain notice of quality guideline violations and file for a site reconsideration.

4.1.2.2 Google Analytics

Analytics are software programs that produce metrics. Metrics are measurements which can help you benchmark desired results. Google analytics take raw data (collected using a small snippet of JavaScript code) and produce information out of that data which is understandable and can be used to improve a website's performance. (Ledford, 2009) Hence Google analytics tool is used to monitor web traffic and to measure the website online performance. By installing Google analytics tool following functionalities can be achieved;

- List of keywords/search query used by visitors to reach your site with geographical, internet and page views information.
- It helps to find the most visited content of your website and pages.
- It indicates a term bounce rate which describes the time spent and deep links of your website visited by users.
- It allows user to set goal on specific page a website's owner wants.
- It offers to analyze visitor trends, loyalty, browser capabilities and network properties.

4.1.3   Really Simple Syndication (RSS)

Most social networking sites offer their content to be syndicated via two main standards which are Atom or RSS (Really Simple Syndication). Feeds include entries such as headline, summaries, excerpts, articles, links to Web content and XML tags. RSS automatically informs internet users about new updates of their favorite websites or blogs. Regular internet users easily stay informed about latest information from the sites they interested in. Some search engines index personnel home pages faster than other pages so it is better to store RSS on iGoogle.com, my.Yahoo.com and my.MSN.com. (Blankson, 2008) Submitting feeds to top RSS directories and sites like feed burner, blog-connection, rss-specifications, rss-submit can improve your website SEO effectively.

Using RSS you can quickly and easily add fresh content to your website. Article feeds are updated regularly, so you can give your visitors and search engines fresh content. You can use RSS to promote any new content, such as new products, special offers, articles and reviews. (Enaceanu, 2007)

4.1.4   Search Engine Ranking factors

Following are the search engine ranking factors identified by three individual entities i-e SEOmoz, Independent Survey and IDV. Table 9 is showing SEO ranking factors comparison in term of their preference (in order of numbers) and usage in SEO. For example, SEOmoz gives most preference to keyword use in title tag whereas Idv thinks that keyword in URL is most important.

#	SEOmoz '07	Survey	Idv
1	Keyword use in title tag	Keywords in title	Keyword in URL
2	Anchor text of inbound link	Keywords in domain name	Keyword in domain name
3	Global link popularity of site	Anchor text of inbound	links Keyword in title tag
4	Age of site	Keyword in Heading Tag	Keyword in H1, H2 and H3
5	Link popularity within the sites internal link structure	Keywords in URL	Page Rank
6	Topical relevance of inbound links to site	Anchor text from within the site	Anchor text
7	Link popularity of site in topical community	Internal links	Site listed in DMOZ Directory
8	Keyword use in body text	Keywords in Alt attribute of images	Site listed in Yahoo Directory
9	Global link popularity of linking site	Relevance of external links	Rank Manipulation by Competitor Attack
10	Topical relationship of linking page	Keywords in body	Site Age

Table 6 Various Ranking Features opinions from Different SEOs (Ao-Jan , 2010)

Where,
SEOmoz'07 is a list of top 10 ranking factors created by surveying 37 SEO experts by SEOmoz (Seomoz.org) in 2007. The Third column labeled by *Survey* is a list of top 10 ranking features rated by a poll of Internet users interested in this topic. This column represents the perception of Google ranking algorithm from general Internet users. The fourth column labeled as Idv is the top 10 ranking features list posted by an Internet marketing expert from idv (data visualize Software Company) on his personal web page. This column represents an individual investigator that studies this topic.
.

4.1.5 Search Engine Marketing (SEM)

The core aim of SEM is to expose the website to potential visitors/clients and attract them, so they visit the website and either use a service or buy a product. SEM is a marketing tool which helps to understand the marketing strategy in the online (virtual) world. SEM consists of many techniques that can be used to boost a product and brand visibility. However Search Engine Optimization (SEO) and Pay per Click (PPC) campaigns prove to

be the most effective techniques for generating high ROI. ROI (Return on Investment) is the key metric when comparing PPC and SEO, as the cost of investment can vary between the two. ROI is calculated as a difference of "gain from investment" and a "cost of investment", divided by "cost of investment". (Adam, 2010)

Google's pay per click (PPC) Adwords model consists of allowing businesses to choose keywords and generate adverts that relate to their site and bid (against competitors) how much per click their business is eager to spend in order to generate traffic direct to their site. The principle behind the bidding process for keywords is simple, if you are willing to pay for the position more than your competitors, your link will be displayed above your competitors. Therefore the additional popular and the higher demand of the keyword, the more expensive a single click is. Adwords offers businesses an immediate opportunity to compete and target new audiences in different or new geographical areas, giving a greater reach and control over their site, depending on a budget set. (Adam, 2010)

4.2   Google Panda Update

The Panda update is Google's newest change in how its algorithms find out the best pages to present in the SERPs in response to search queries. Google's goal is to create the search experience and browsing as smooth and as fast as flipping through a magazine. The Panda updates are an attempt to attain the objectives of quality search results and speed. The main focus is on quality content and the usability of the site for the user. Websites and blogs with high-quality and interesting content will get rewarded, as will those that offer a fluid user experience with streamlined navigations and easy-to-find content. About "quality content" Google has provided a list of questions at this link (http://googlewebmastercentral.blogspot.com/2011/05/more-guidance-on-building-high-quality.html) where you can inquire yourself as you think developing content for your site. (Jones, 2011)

Some important points related to content development are (Jones, 2011)

Written by expert: Content should be written by expert copywriters who know and understand the subject well.

Keyword variations: Avoid duplicate the content for multiple pages and then switch out different keyword. You should make sure each page is uniquely written around the main topic of the page.

Generate interest of reader: Each page or article/content should map to the genuine interest of your readers and visitors. When you examine keywords with social media tools, you can place into the proper context by watching how they are used in a sentence or a conversation.

Good Value: As you consider the content you plan to create, you should find out what has already been written on the subject? How does your written article compete to others that have been written? Is it unique and how much detail in it?
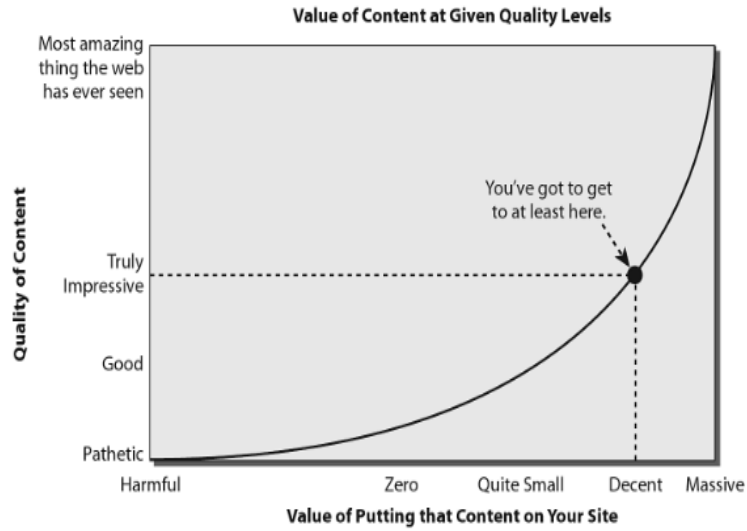
Figure 7 Value of the content at given quality levels (Jones, 2011)

Rand Fishkin, cofounder of SEOmoz.org and a well known in search industry has posted a chart to illustrate this principle and outlines where you require being as you consider putting content on your website as described in figure 7.

5   BLACK-HAT SEO

Black Hat SEO implements malicious website ranking optimization techniques which violates the search engine optimization rules. Implementing black hat techniques may result in a short term rise in organic listings. However, if it is discovered by the search engines that websites who use black hat techniques then those sites may be penalized (by demoting it further down the rankings) or removal from its index completely. Being penalized will obviously have a major impact on a site's reputation and performance in the future. Black hat SEO techniques include using hidden words and links (invisible to the user) in attempt to deceive search engines. Keyword stuffing consists of over loading a web page with keywords or irrelevant keywords in order to manipulate a site ranking. (Adam, 2010)

The main Black-Hat SEO methods are as follows: (Wu Di, 2010)

5.1   Doorway Pages

It uses software to generate automatically a large number of pages including keywords and then automatically turn these pages to the home page. And the purpose is to let these Doorway Pages with different keywords get good search engine rankings. (Wu Di, 2010)
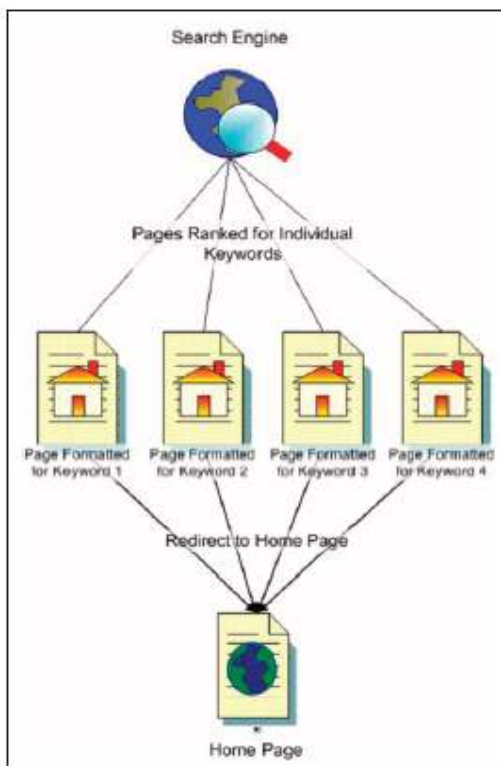

Figure 8 (Example of Door way pages working (Malaga, 2008))

The goal of the doorway pages to gain high rankings for multiple keywords or terms. The optimizer will create a separate page for each keyword or term. Some optimizers utilize hundreds of these pages. Doorway pages typically use a fast Meta refresh to redirect users to the main pages as explained in figure 7. A Meta refresh is an HTML command that automatically switches the users to another page after a specific period of time. Meta refresh is generally used on out of date web pages. For example, you might see a page that states" you will be taken to a new page in 5 seconds." (Malaga, 2008)

5.2 Keyword Stuffing

Stuffing large pile of keywords in the page those causes increase in keyword density of the page. The keyword density is factor which can increase page relevance for keywords. (Wu Di, 2010) This means, adding too much keyword in a single paragraph to make that keyword importance for search engine.

5.3 Hidden Text

The keywords are put into the HTML files. However, these words cannot be seen by users or naked eyes, but search engine can only recognize. (Wu Di, 2010) For example, the text presented with the same colour scheme as the corresponding web page has.

5.4 Hidden Link

Put the keywords into the link, and this link is invisible to the user. (Wu Di, 2010)

5.5 Cloaked Page (Cloaking)

Programs or scripts are used to detect whether the access is the search engine or a normal user. If it is search engine, there will return optimized web pages. If it is the common people, the return will be another version. (Wu Di, 2010) The purpose of cloaking is to get high rankings on all of the major search engines. Since each search engine uses a different ranking algorithm, a page that ranks well on one may not necessarily rank well on the others. Since users will not see a cloaked page, it can contain only optimized text, as no design elements are needed. So the black hat optimizer will set up a normal Web site and individual, text only, pages for each of the search engines. The final step is to examine requesting IP addresses. Since the IP addresses for most of the major search engine spiders are well known, the optimizer can serve the appropriate page to the correct spider (see figure 8). (Malaga, 2008)
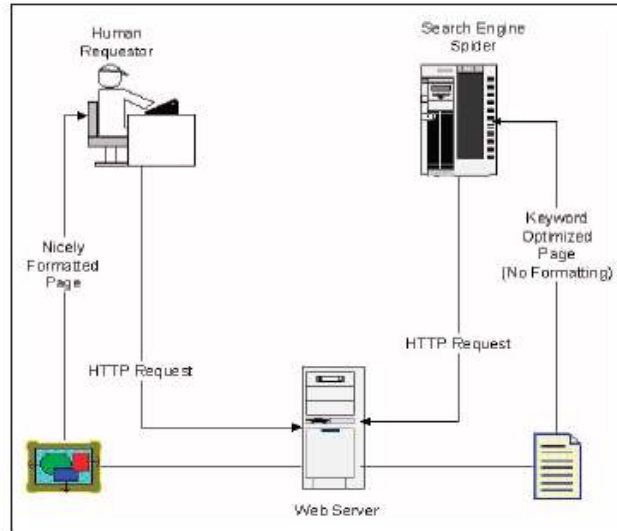
Figure 9 (Cloaking working example (Malaga, 2008))

5.6   Content Generator

It is practice by Software that searches the Web for specified content and then copies that content into a new Web page. These so-called content generators in fact search the Web for keywords and terms specified by the optimizer. The software then mainly copies content from other sites and includes it in the new one. Content generators represent a problem for legitimate Web owners as their original content may be copied extensively. Since some search engines penalize duplicate content, legitimate sites may also be penalized. (Malaga, 2008)

5.7   HTML Injection

HTML injection occurs when a user utilizes security vulnerability in Web site search programs, by sending the program a search string which contains special HTML characters. These characters source the insertion of data specified by the user into the site. It allows optimizers to include a link in search programs that run on another site. For example, WebGlimpse is a Web site search program widely used on academia and government Web sites. The Stanford Encyclopedia of Philosophy Web site located at plato.stanford.edu, which has a Google page rank of 8 (links from sites with a high page rank are highly valued), uses the WebGlimpse package. So an optimizer that would like a link from this authority site could simply navigate to
http://plato.stanford.edu/cgi-
bin/webglimpse.cgi?nonascii=on&query=%22%3E%3Ca+href%3Dhttp%3A%2F
%2F##site##%3E##word##%3C%2Fa%3E&rankby=DEFAULT&errors=0
&maxfiles=50&maxlines=30&maxchar s=10000&ID=1.

The optimizer then replaces ##site## with the target site's URL and ##word## with the anchor text. (Malaga, 2008) In above example attacker first get the URL code from webserver which is venerable enough for HTML injection. Attacker edits the address with target keyword and his site address (in above example replaces "site" with its target

website URL and "word" with his target keyword). When user clicks or refreshes the malicious URL then JavaScript or Vb script code run with the privileges of victim user.

5.8 Blog-ping (BP)

It is method for attracting search engine spiders that involves the making of hundreds (or even thousands) of blogs and then continuously pinging blog servers (telling the servers that the blog has been updated). (Malaga, 2008) For example, web page is also submitted to number of websites available like pingomatic.com. It offers the practitioners to inform search engines automatically about the websites or blogs latest updates and changes in web pages. It helps the search engine spiders and crawler to index new created pages faster.

6 OVERVIEW OF SEARCH ENGINE INDEXING AND WORKING

Search engines crawl and index billions of web pages daily. Indexing refers to the process when search engines include web pages in their database to facilitate fast and accurate information retrieval process.

Exploring the content of web pages for automatic indexing is of fundamental importance for proficient e-commerce and other applications of the Web. It enables users, including customers and businesses, to establish the best sources for their needs. Majority of search engines today use one of two approaches for indexing web pages. In the first approach, the search engine selects the terms indexing a web page by analyzing the frequency of the words (after filtering out common or meaningless words) appearing in the whole or a part of the text of the target web page. Typically, only a title, an abstract or the first 300 words or so are analyzed. The second method relies on sophisticated algorithm that takes into account associations of words in the indexed web page. In both cases only words appearing in the web page in question are used in analysis. Often, to increase relevance of the selected terms to the potential searches, the indexing is refined by human processing. (Szymanski, 2001)

6.1 Web Crawling

A web crawler forms an integral part of any search engine. The basic task of a crawler is to fetch pages, parse them to get some more URLs, and then fetch these URLs to get even more URLs. In this process crawler can also log these pages or perform several other operations on pages fetched according to the requirements of the search engines. The entire popular search engines use crawlers that must scale up to substantial portions of the web, however due to the competitive nature of the search engine business, the designs of these crawlers have not publically described. (Zhang, 2005)

The Google search engine is a distributed system that uses multiple machines for crawling. The crawler consists of five functional components running in different processes. A URL server process reads URLs out of a file and forward them to multiple crawler processes. Each crawler process runs on a different machine, is single threaded, and uses asynchronous I/O to fetch data up to 100 web servers in parallel. The crawlers transmit downloaded pages to a single store server process, which compresses the pages and stores them to disk. The pages are then read back from disk by an indexer process, which extracts links from HTML pages and save them to a different disk file. A URL resolver process reads the link file, derelativizes the URLs contained therein, and saves the absolute URLs to the disk file that is read by the URL server. (Zhang, 2005)

In general, three to four crawler machines are used. The internet archive also uses multiple machines to crawl the web. Each crawler process is assigned up to 64 sites to crawl, and no site is assigned to more than one crawler. Each single-threaded crawler process reads a list of seed URLs for it assigned sites from disk into per-site queues, and then uses asynchronous I/O to fetch pages from these queues in parallel. Once a page is downloaded, the crawler extracts the links contained in it. If a link refers to the site of the page it was

contained in, it is added to the appropriate site queue; otherwise it is logged to disk. Periodically, a batch process merges these logged "cross-site" URLs into the site specific seed sets, filtering out duplicates in the process. (Zhang, 2005)

6.2   Web Caching

The larger search engines in particular have to be able to process tens of thousands of queries per second on tens of billions of documents, making query throughput a vital issue. As a result, query processing is a major performance bottleneck and cost factor in current search engines. To satisfy this heavy workload, search engines use a variety of performance optimizations including index compression, caching, and early termination. (Sarina, 2008)

Web caching is the short-term storage of Web objects (such as HTML documents) for later retrieval. There are three major advantages to Web caching: reduced bandwidth consumption (fewer requests and responses that need to go over the network), reduced server load (fewer requests for a server to handle), and reduced latency (since responses for cached requests are available instantly, and closer to the client being served). Together, they make the Web less expensive and better performing. (Sarina, 2008)

Search engine cache the web sites pages and then store into database to show results for user search queries. Search engines take snapshots which include the content of the web pages and Meta tags information with standard HTML format. Google and other search engines cannot cache or read graphic (image, flash) but the content used for these graphics. Google cache websites on regular basis depends upon the website quality, ranking and updating. Huge and lively websites having high traffic and content updating crawl and cached regularly (sometime hundreds of times in a day) by Google. On the other hand Google seldom in a month crawl and takes cache of the website which has poor ranking and no updates.

To assure throughput and fast response times, current large search engines are based on huge clusters of hundreds or thousands of servers, which is dependable for searching a subset of the web pages. To optimize the efficiency of the query result large result caches are employed and a portion of the query traffic is served using formerly computed results. In web search engines, cache components happen at several levels and in various forms, e.g., result, posting list, and document caches. The result cache which we focus on may be deployed in divide or the same machines with query processors. (Hui, 2010)

At the lower level of the system, index structures of frequently utilized query terms are cached in main memory to save on disk transfers. At the higher level, a search engine obtains a query from a user, processor the query over its indexed documents, and returns a small set of related results to the user. If an earlier computed set of results is cached, the query can be served directly from the cache without executing the same queries over and over again. (Hui, 2010)

6.3   Web Indexing & Searching

Internet users perform billions of queries on web search engine daily. Most of common search engines like Yahoo, Google and Bing fetch results in milliseconds for the particular query, however it has been seen that results from the search queries vary (different) in different search engines respectively. The ability of web search engine to generate fast and millions of websites in short time depends upon number of factors. Web searching involves several processes through crawling to display the URLs results.

Each web search engine depends upon one or more crawlers to present the content for its operation. Crawlers use a starting set of uniform resource locators (URL). The crawler retrieves (i-e copies and stores) the content on the sites specified by the URLs. The crawlers extract URLs appearing in the retrieved pages and visit some or all of these URLs thus repeating the retrieval process. Each search engine follows its own unique timetable for re-crawling the web and updating its content collection. The web search engines store the web content they retrieve during the crawling process in the page repository. (Amanda, 2004)

The indexer processes the pages crawled by the crawler. It first chooses which pages to index, for example, it might discard duplicate documents. Then, it makes various auxiliary data structures. Most search engines build some variant of an inverted index data structure for words (text index) and links (structure index). The inverted index contains for each word a sorted list of couples (such as docID and position in the document). (Pokorny, 2004)

The result is generally a very huge amount of database provide the URLs that point to pages where a particular word occurs. In this area, search engines have much in common with traditional IR (information retrieval) systems in terms of techniques they utilize to organize their content. Given the hypermedia aspects of much of the Web content, however, the database may also contain other structural information such as links among documents, incoming URLs to these documents, formatting aspects of the documents, and location of terms with respect to other terms. (Amanda, 2004)

The Web query engine obtains the search requests from users. It gets the query submitted by the user, splits the query into terms, and searches the database made by indexer to locate the terms and hence the Web documents referred by the stored URLs. Then it retrieves the documents that match the terms within the query engine then retrieves the documents that match the terms within the query and revisits these documents to the user. The user can then click on one or more of the URLs of the accessible Web documents. (Amanda, 2004) This process is shown in following figure 10.
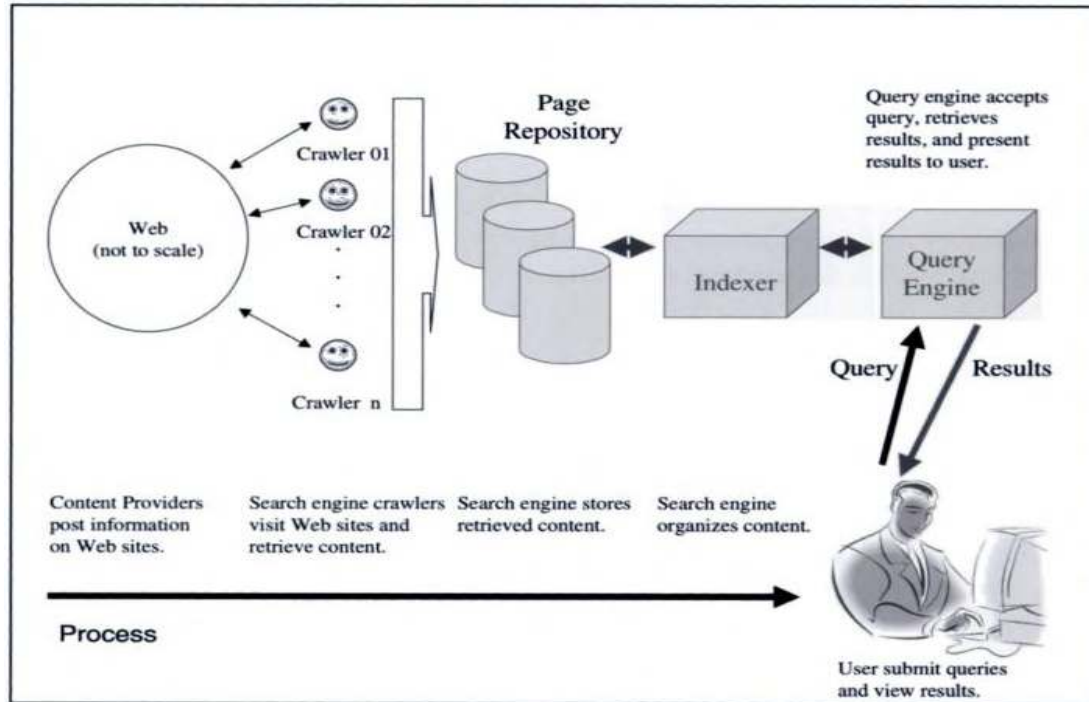
Figure 10 (Basic Web Search Engine Architecture and Process (Amanda, 2004))

6.4   Methods of Ranking Documents (URLs)

The basic approach to ranking and matching used by Web search engines is basically to take the terms in the query and place items that hold at least some terms. There are a lot of variations on this basic approach. Some of techniques generally used in matching and ranking algorithms consist of: (Amanda, 2004)

6.4.1   Click Through Analysis

Click though analysis utilizes data concerning the frequency with which users chose a particular page as a means of future ranking. Click through analysis consists of logging queries and the URLs searchers visit for those queries. The URLs that are visited most by searchers visit for those queries. The visited URLs most often by searchers in response to a particular query or term are ranked higher in future results listings. (Amanda, 2004)

6.4.2   Link popularity

Websites with huge amount of back links generally have higher crawling rates. This is because due to enormous link popularity the website gets good PageRank as well as ranking on many keywords and content which ultimately boost the click through analysis as well. For example, consider site A ranked with a few number of ordinary keywords compare with a huge website site B with hundred of keywords or ranked on high search volume keywords. Because of high ranking site B get great amount of users which ultimately increase the click through rate as well.

6.4.3 Term Frequency

Term frequency is a numerical evaluation of how frequently a term appears in a web document. Normally, a greater frequency of occurrence tells greater chances that the document is related to the query. In general, and within special domains such as law, documents use a small subset of terms relatively and a great subset of terms very infrequently. Web search engine may preserve a stop list of such words, which are excluded from both indexing and searching. (Amanda, 2004)

6.4.4 Term Location

Term location many times indicates its importance to the document. Therefore, most Web search engines give more weight to certain terms (for example, those occurring in Title, lead (first) paragraphs, image captions, special formatting such as bold, italic etc.) than terms occurring in the body of the document or even in a footnote. (Amanda, 2004) It means while ranking documents search engines give more importance to web page title, the first paragraph in the content.

6.4.5 Term Proximity

Term proximity is the distance between two or more query terms within a document. The rationale is that it is more likely that the document is relevant to the query if the terms appear near each other. Mostly web queries are short (i-e contains one or two terms), term proximity is various times of little value. However, in definite situations, term proximity can be great value. Name searching is one such area. (Amanda, 2004) For example, a search term could be used to find "buy camera Finland" which match phrases such as "Buy camera in Finland" or "camera buy in Finland". By keep limiting the proximity, these phrases can be matched while evading documents where the words are speckled or widen across a page or in not related articles in a paragraph.

6.4.6 Text Formatting

Google and other search engines generally follow the standard HTML format. Major search engines cannot read graphics, images and flash data directly but only the text use in form of alt tags, embedded tags for images and flash respectively.

Text formatting is the use of specific HTML code in their ranking algorithms. Even though several schemes (e.g. bold terms, emphasis tags, etc.) the most successful of these is the use of anchor tags. When an information provider produces a Web Page, the document will a lot of times contain links to other Web documents or links to particular locations in the documents. Readers of the documents see these all links as clickable text which is called anchor text. Anchor texts and tags are valuable resources for search engines. (Amanda, 2004)

7 INTRODUCTION TO CONTENT MANAGEMENT SYSTEMS

Content in general refers to any kind of textual information, visual, sound, audiovisual or a meaningful data. Content can be produced, archived, altered, spread, consumed and published in parts or entirely. Content therefore, is information that you tag with data so that computer can manage and systematize its collection, management, and publishing. (Boiko, 2005)

7.1 What is Content management system

Content management system is a broad term which covers a collection of processes and procedures to organize work flows in a shared environment. Content management is a set of procedures and technologies that manage organize and publish information in any form, shape or medium. Content management has an important impact in managing organizational knowledge and work flows today. A content management system (CMS) is used to manage, store, organize all the text and images of your website. (Lopuck, 2006) This refers more to a web content management system. Content management system deals with the creation of the content, managing, distributing, publishing, storing and finding the information and data. A web content management system (WCMS) covers the whole lifecycle of a website from providing the facility to create, manipulate and manage the content through to publishing content and also to archiving. The other key functionality of web content management system is to manage the structure of the website, appearance of the content, tags, published pages and the navigations of the website.

Figure 11 (Basic CMS functionality)

A content management offers user to find useful, relevant and reliable information within and outside the organization. An ideal content management system will have following properties (Farida Hasanali, 2003)

- Set of standardized technologies to manage content in various formats.
- CMS has a taxonomy that is customer-driven and domain-driven.
- It provides a work flow to move and manipulate content efficiently through the content validation and approval process.

- It has a metadata layer that offers information about the content in addition to the taxonomy.
- It allows front-end applications (e.g. portals, blogs) to present meaningful information.
- It tracks how information and other data are used.

7.2 Features of Content Management Systems

A well constructed content management system has some key features which are centralized repository, rapid content import, work flow automations, authorization and security, dynamic tracking and alerts, version control, automatic distribution and publishing software. Authors using a content management system need functions that control the authorization process and administrative tools. Key features of a web content management system are

- A rich content management system authorization controlled by End users.
- Multi user authoring.
- Navigations Edit and control.
- Styles and Templates with available design and presentation.
- Workflow management
- Setting rights and roles of users and authorization.
- Media storage which includes storing images, visual data like videos, sound files with usability features.
- Ease of use and efficiency.
- Version control, auditing and archiving.
- Reporting and Multi language functionality.
- Security and integrity of the content.
- Usability and accessibility.
- Brower support.
- Time management (speed) for publishing and viewing.

7.3 Web development for CMS

Content management systems are key development trend for small and medium-sized businesses. The ease of web development occupies heavily by CMS invention as now most of the websites on internet are CMS based applications and ECMSs. The features of content maintenance, website structure editing, customization and incremental growth create CMS model and applications attractive for online businesses. It is important that a CMS should be a quality, cost effective and modifiable solution. Ideally the factors in development process of CMS solution are (w3c, 2011)

1. It should be legally compliant because it will go through w3c validation process.
2. It should be standards- compliant according to w3c standards. (w3c, 2011)
3. Should be search engine compliant: according to the search engine standards.
4. It is important that it should be developed visitor/user-friendly.

5. Since search engine don't like so java scripts, frames, heavy code so keeping code light weight and content oriented is ideal for development.
6. It should follow the basic fundamental structure of HTML(e.g. use of meta tags, heading, alt and other tags)

Most of the CMS development now is built at high standards which apply significantly in open source software market to large enterprise content management applications. In most implementations three elements of the CMS vary by product but commonly include (Randy, 2006)

- The compulsory completion of specific fields, such as Meta data.
- The HTML code in which content will be displayed.
- The navigation structure in which content will be embedded.

SEO techniques and tactics use to tune each of these elements such that they appear more search engine friendly and to their ranking algorithms.

7.4 Web 2.0 Integration

Web 2.0 is a set of economic, social, and technology trends that jointly form the foundation for the next generation of the Internet. Moreover, it is a full-grown distinctive medium characterized by user participation, openness, and network effects. (Musser, 2006) Web 2.0 is more about use for internet with latest updates and is described by interactive applications which allow user to organize, distribute, create and contribute their content. The quality of sites after integrating web 2.0 gets better and improves search engine rankings. Popular examples of Web 2.0 are YouTube, flicker, Facebook, Hi5, MySpace, Wikis, blogs and many other social networking websites. Evolution of Web 2.0 offers lot of advantages over Web 1.0 like two way communication, peer to peer, dynamic, XML integration, active and collaborative features. Comparison of Web 1.0 and Web 2.0 is shown in the Figure 13.
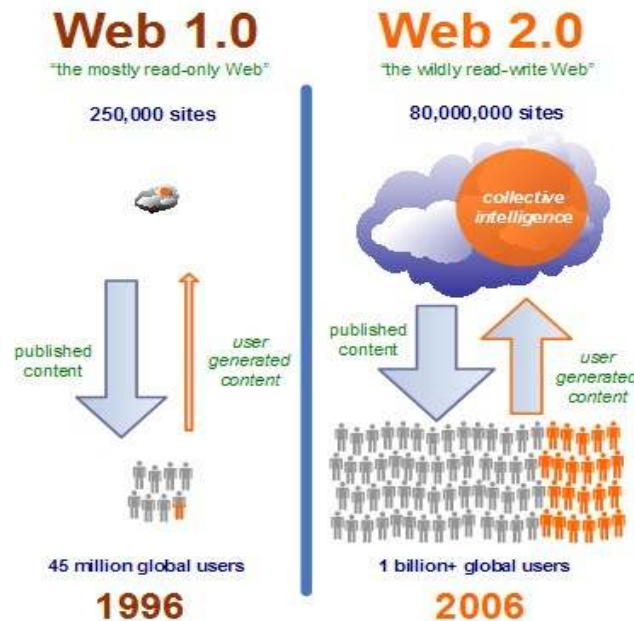
Figure 12 (Comparison of Web 1.0 and Web 2.0 (portfolios, 2008))

Web 2.0 websites allow users not only to retrieve information rather it is a platform which facilitates users to collaborate, share information and ideas. Web 2.0 when integrates with content management systems then it provides user blogging, podcasts, RSS feeds, online communities with development features like interactivity, usability and adjustment of latest updates. Using Web 2.0 it is easy to filter traffic, interact with other 2.0 applications, for example linking to web 2.0 sites worth more than other websites. Web 2.0 development has many advantages which are as follows: (Graham, 2008)

- Users as first class entities in the system, with well-known profile pages which includes such features as: age, sex, location, testimonials, or comments about the user by other users.

- The ability to form connections between users, via links to other users who are "friends", have membership in "groups" of a variety of kinds, and subscriptions or RSS feeds of "updates" from other.

- The ability to post content in lot of forms: photos, videos, blogs, comments and ratings on other users' content, tagging of own or others' content, and some skill to control privacy and sharing.

- Other more technical features, including a public API to agree to third-party enhancements and "mash-ups", and embedding of different rich content types (e.g. Flash videos), and communication with other users through inside email or IM systems.

50

- It offers a lot of features in a single platform like post photos, guest books, views exchanges, maps, tags, communications. (Graham, 2008)

Feature class	Feature	Facebook	YouTube	Flickr	LiveJournal	MySpace	Digg	Friendster	Amazon	Ebay	Craigslist	Slashdot
Profile details	Age	✓	✓		✓	✓		✓				
	Location	✓	✓		✓	✓	✓	✓	✓	✓		
	Gender	✓				✓		✓				
	Testimonials	✓	✓	✓		✓		✓		✓		
	Other data	✓	✓	✓	✓	✓	✓	✓				✓
Connectivity	Friends	✓	✓	✓	✓	✓	✓	✓	✓			✓
	Subscriptions		✓	✓	✓			✓				
	Groups	✓		✓	✓	✓		✓				
Content	Main content	profiles	videos	photos	blogs	profiles	links	profiles	products	products	ads	articles
	Other content	photos			(photos)	blogs, video photos			photos	photos	photos	
	Tagging	✓	✓+	✓	✓+				✓			
	Friends only	✓		✓	✓	✓						
	Comments	✓	✓	✓	✓	✓	✓		✓	✓		✓
	Editable content								✓			
	Rateable content		✓	✓				✓	✓	✓		✓
	Viewing Statistics		✓	✓				✓		✓		
Technical	Public API	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	Embedding allowed	✓				✓	✓					
	Many RSS feeds		✓	✓	✓	✓	✓				✓	
	Private messages	✓	✓	✓	✓	✓		✓				

+ Only content creators are allowed to assign tags.

Table 7 (WEB 2.0 features in some of most popular websites on internet (Graham, 2008))

Open source market offers a lot of content management systems with distinct features. According to CMS intelligent report 2010 Drupal, Joomla and WordPress are the top CMS systems in today's market which is based on most widely used and evaluated CMS worldwide. (Interactive, 2010) The main purpose is to study the features of each of CMS system in term of search engine optimization and effective development. Hence to analyze which CMS scores better regarding SEO and for which type of application, it is suitable.

7.5   Most Common Content Management Systems & Features

There are enormous amount of CMS available on the internet. The most popular CMS systems are those which are open source and free of cost. Today, most commonly use CMS systems are WordPress, Joomla and Drupal and occupying most of CMS market share.

According to Alexa's rankings for the top one million websites run on which CMS in terms of the technologies used on those sites WordPress get over 12.5% of site under it.

Joomla is on second place with almost 2.5% share, and Drupal goes down to 3rd with 1.4%. (Savan, 2011)

7.5.1 Drupal Content Management System

Drupal is a free open source software package that offers anyone to easily publish, manage, control and organize a wide variety of content on a website. Hundreds of thousands of people and organizations are using Drupal to power a continuous variety of sites. (Drupal O. , 2005)
Drupal is easy to use, based on modules, extensible and standard-based content management system. Drupal modules and power pages offer much functionality such as blogs, picture galleries, extendible content, forums, community portals and corporate e-commerce sites.

Drupal allows system administrators to (a) create and organize their own site content, (b) automate administrative tasks, (c) manage site contributors and visitors, and (d) customize site presentation. Drupal has known to be search engine friendly. It has a clean and standards-compliant markup and builds all the tags that the search engine is looking for during the "search process". It also offers users to obtain complete control on the path of the page as the creation of short yet clean and informative paths will definitely improve a site's search page ranking. (Ryals, 2008)

7.5.1.1 Drupal SEO Modules and Utilities

Drupal offers extensive features considering search engine optimization ranking. Important features I found are as follow;

i. Drupal provide advanced URL control system having nice clean URLs. It also provides the facility of URL alias which can be given for each item of content.

ii. Pathauto module automatically produces path aliases for various kinds of content (nodes, categories, users) without requiring the user to manually specify the path alias. Path redirect, Global redirect, Pathologic, Sub-path URL Aliasing modules provide search engine friendly URL structure. (Drupal, 2011)

iii. Drupal provides Content Construction Kit (CCK) and view modules to generate advanced custom views and new content types. Examples are blog posts, classified ads, podcasts, and news stories.

iv. Users can configure Drupal to save a new version of website pages every time they are editing means user can go back to see the old version. Hence Drupal gives you revision control facility. (Drupal W. , 2011),

v. Drupal provides a powerful taxonomy (category) system that offers user to organize and tag content. Drupal advanced taxonomy features, combined with CCK and Views, allows you to effortlessly target all of those long-tail keywords that you researched in keyword research (e.g. Wordtracker).

vi. Drupal was designed for community-based Web sites and provides strong user role and access rule functionality.

vii. Drupal's Page Title Modules gives user custom control of HTML elements, while the Meta Tags Module offers control over webpage's individual meta description tags. (Drupal W. , 2011) This is very important regarding SEO point of view because every individual page Title and Meta tags should be unique according to page content.

viii. Drupal provides documentation that includes the official handbooks, massive API Reference, frequent tutorials, videos, blogs and podcasts and new book Pro Drupal development. (Drupal W. , 2011)

ix. *The SEO Checklist module* (http://drupal.org/project/seo_checklist) provides a checklist of SEO best practices to help you to maintain track of what needs to be done on your website. It checks to see all modules installed on your website and you a list of "to-do" items. It also offers you with a date stamp next to each item after it marked as completed. This is a great module to keep track of all SEO-related tasks that's need to be completed on a website, particularly if multiple people are working on a website. (Townsend, 2010)

x. Drupal provides a various amount of modules for SEO mentioned in figure 14.

SEO Checklist	SEO compliance Checker	Pathauto	Path Redirect	Global Redirect
Search404	Meta Tags + Nodewords	Global Geourl	HTML Purifier	Page Title
Menu Attributes	New XML Sitemap	Site Map	Token	Nofollow List
WYSIWYG API	FCKE Editor	IMCE	Mollum/Spam	Captcha
Actions	Triggers	Notify	Scheduler	
ADDthis / Diggthis/ Sharethis	Advanced: Apache Soir Search)	Advances: Open Calais-RDF metadata		Google Analytics

(In Red Font) Modules that directly affect SEO---- Must have for SEO
(In Black Font) Modules that indirectly affect SEO--- Should have

Table 8 (Drupal Modules for SEO **(Forgan, 2009)**

7.5.2 WordPress Content Management System

WordPress is one of many PHP/MySQL content management systems that offer content editors to use a web interface to maintain their sites instead of editing and uploading

HTML files to server. WordPress started as a blogging tool, but early on the developers added pages as a separate content type which offers an easy, web interface to create and manage web content. Despite winning the Overall Best Open Source CMS Awards, WordPress is still extensively considered be a blogging tool. WordPress is famous for its five-minute installation and comes with everything to setup a basic website. The core system includes: (Leary, 2010),

Post and Pages: A blog (composed of posts) will feature a few "static" (database-driven) pages, such as "About", hence these are two primary content types in number of other ways.

Media files: The post and pages editing screens offer you to upload images, audio, videos, office documents, PDFs, and more.

Links: WordPress includes a link directory, often referred to as the blogroll.

Categories and tags: WordPress consists of both hierarchical and free-from taxonomies for posts. There is a separate set of categories for links.

User roles and profiles: WordPress users have five possible roles with growing capabilities (Subscriber, Author, Contributor, Editor, and Administrator) and a very basic workflow for additional approval. User profiles include a description, avatar and several forms of contact information

RSS, Atom and OPML feeds: There are RSS and Atom feeds available for just about everything in WordPress. The major feeds include recent posts and comments, but there are also feeds for individual categories, tags, authors and comment threads. An OPML feed for links is also built in.

Clean URLS: With the included .htaccess files, WordPress supports search engine friendly URLs (or permalinks) on both Apache and IIS servers, with a system of tags that allow you to customize the link structure and several built-in configurations.

Spam Protection: The WordPress include Akismet plugin, which provides industrial-strength filtering of spam comment.

7.5.2.1 WordPress SEO Modules and Plugins

There are many WordPress SEO plugins are available from them most powerful and widely used is WordPress "SEO Ultimate plugin" and "All in One SEO pack" because it provides many SEO utilities in one package. SEO ultimate plugin provides control over title tags, noindex, Meta tags, canonical[3] URLs, autolinks, 404 errors and more.

The ultimate plugin list offers maximum benefits for traffic and search engine friendly blogs. "All in one SEO pack" plugin provides following SEO features (Torbert, 2011)

- It provides Google Analytics support
- Give support for Custom Post Types
- Advanced Canonical URLs
- Built-in API so other plugins/themes can access and extend functionality

---

[3] Canonical URLs are URLs that are considered to be authentic and authoritative for a given resource. The problem with many blogs platforms is that the same content often appears in different places on a site with exactly the same presentation but an entirely different URL. For example, http://example.com and http://www.example.com show same content in web browser. (Brazell, 2010)

- Only plugin to offer SEO Integration for WordPress e-Commerce sites
- Nonce Security
- Support for CMS-style WordPress installations
- Automatically optimizes your web pages' title for search engines
- Generates META tags automatically
- avoid the distinctive duplicate content found on WordPress blogs
- For beginners, you don't even have to look at the options; it works out-of-the-box. Just install.
- For expert users, you can fine-tune everything
- You can override any title and set any META description and any META keywords you want.
- Backward-Compatibility with many other plugins, like Auto Meta, Ultimate Tag Warrior and others. (Torbert, 2011)

7.5.3 Joomla Content Mangement System

Joomla is one of the most famous open-source Content Management System (CMS) platforms. The default installation of Joomla is already a multi-functional content management system, but if the default basic features cannot meet your requirements, you can easily expand it with extensions. There are five types of extensions available for Joomla. These are Components, Modules, Plugins, Templates, and Languages. Each of these extensions deals with unique functionality. (Cao, 2010)

Joomla is popular because of following features. (Cao, 2010)

1. Very easy to install and use with lots of extensions and modules
2. The documentation is very exhaustive and concise
3. Admin user interface is intuitive and powerful
4. The backend of Joomla is very usable and the WYSIWYG editor the content was nice

A number of search engine optimization extensions can help with making your Joomla more search engine friendly. One of the popular extensions is SEOSimple by Dao Design. The extension plugin works by taking the first chunk of text in an article or page and then uses it as the metadata for the description of the page. This helps sites evade duplicate metatags on all their pages such as the main keywords and description that are present in the global configuration of Joomla. (Waring, 2010)

7.5.3.1 Joomla SEO Features

Joomla provides a separate SEO settings widget while installing the website for the first time in super administrator mode. While installing Joomla in the global configuration settings panel there is a panel available with tag "SEO Settings" with two functionalities which are search engine friendly URLs and use mode rewrite. Use mode rewrite allows user to rewrite the website URLs to make them search engine friendly.

When you first install, your URLs will look like
http://www.joomlaapetcenter.com/index.php?option=com_content
&task=view&view=article&id=11
But when you activate SEF URLs, the same URL would look like something like:
http://www.joomlaapetcenter.com/dogs/collies

Keywords and metadata play an important role in identifying your content. Joomla enables you to add globally, to each individual article, or both. Metadata information is stored in the page header: (Crowder, 2009)

Example:
<meta name= "keywords" content ="hosting, web hosting"/>
<meta name= "description" content ="web hosting as low as $10"/>
<meta name= "author" content ="Abrar"/>
<meta name= "robots" content ="index, follow"/>
There are many SEO extensions used for Joomla from them AceSEF and sh404SEF are most widely used among others. Sh404SEF use for URL rewrites which can be downloaded accessing this URL (http://extensions.joomla.org/extensions/site-management/sef/10134).

8   FINDINGS AND RECOMENDATIONS

8.1   Discussion

Search engine optimization is a long term process but with great benefits in terms of lead and revenue. On the other hand Search engine marketing is a limited term process depends upon the budget spends on the campaigns and there is a high risk involve to get ROI and to generate high revenues. White-hat SEO techniques are widely practiced by most of SEO practitioners while very few use Black-hat SEO because by using Black-hat SEO website can be penalized or banned by major search engines especially Google. Once a website is banned cannot be recovered in the same position before it has.

Search engine optimization and marketing becomes an essential need for every online and even offline business. In this era of internet everyone wants to shop and get services sitting at home. SEO industry is grown hugely in the last decade and big giants of internet like Facebook, YouTube, Doubleclick etc. Search engine techniques are continually updating with the change and updating of ranking and search algorithms of major search engines. For example, in 2010, Google made a major change in its algorithms which names as Google Panda. Introduction of Google panda made a huge impact in SEO strategies and techniques, now SEO practitioners are more focusing on content creditability, GUI and structure of the web. With the introduction of Google Panda, SEO is become more ethical rather working on traditional old SEO methods.

Invention of Content management system and open source PHP has made the web development easier, efficient and cost effective. There are hundreds of open source CMS applications available on internet from them major contributors are Joomla, WordPress, Drupal, PHPNuke, IBM Enterprise, Mambo etc. Most of known CMS are not designed for SEO that's why developers need to change in code and apply additional plugins to make websites more SEO friendly. Web 2.0 with content management systems has made a great impact in term of making website with more features, online, search engine friendly and application integration like RSS, tagging, API easier.

It is very important to know for search engine practitioners how search engine works in backend, the processes in fetching data and executing queries. Search engine algorithms update with passage of time nobody knows what exactly they have changed, and what are their updated ranking criteria. In 2010, Google search engine made a major change in their ranking and caching algorithms named as Google Panda. Google Panda has changed the SEO strategies dramatically with implementation of more ethical and user oriented SEO practices rather traditional SEO.

8.2   SEO Outcomes & Recommendations

Search engine optimization is set of processes, tools and techniques continuously during the project. It is very important to keep update yourself about latest search engine ranking and algorithms news and their updated criteria.

The SEO process can be divided into three main parts working in cycle, which are Initial Analysis and SEO Goal, On-Page SEO and Off-Page SEO. Before starting a SEO project it

is important to analyze business domain and industry you are targeting. At very first step choose a best domain name for online business. For example, if your business is of hosting, it is better to have "hosting" keyword in the domain name. Secondly important thing is to purchase or host the website with dedicated IP address. Uploading the website with shared hosting and shared IP is very less beneficial than hosting dedicated IP. Google and other search engine use IP of website as primary address. Since there can be bundle of websites from which some sites can be panelized or disliked by search engines can be shared on shared website that's why it is always a good practice to purchase a dedicated IP if you want to grow your business and perform better.

After you have chosen the domain name it is time to do SEO analysis and set SEO goals and strategy. In SEO goals and SEO strategy phase you set your SEO campaign targets, target area, target audience and budget. In SEO analysis the first thing to do is the keyword research where you identify most suitable keywords for your SEO campaign. To determine the best keywords keyword effectiveness index and keyword opportunity index is calculated.

For example, A business with printing cards services can choose keywords use in above Table 7.
From table 7 sample keywords No1 and 2 clearly have the maximum KEI and KOI which means to get ranking on these keywords will able to get maximum web traffic from search engines.

8.2.1 Competitor analysis

Competitor analysis is very important before start working on the website and make changes. During competitor analysis following outcomes help SEO practitioners in their SEO campaigns.

- Search engine indexing and visibility of the competitor websites.
- Page rank and outbound links (their partner's websites) they are using.
- Their targeted keywords and on which them they are ranked well in search engines.
- An estimated value of web traffic competitor websites is getting. (Using Google Doubleclick tool)
- The back links of the competitor websites. However it is very difficult to find all back links for a competitor website but using SEO tools and search commands (link:address or link: website address) in search engines you can get which are included in Google cache and other search engines. It is very important to know what kind of link popularity strategy your competitors are using.
- It helps to get an idea about content strategy and competitor website's structure.

8.2.2 Website Code and structure

Following are the key recommendation while applying SEO techniques on the website code and structure.

Title Tag

- Title tag must include a keyword phrase with 2-3 words combination. The ideal length should be between 55-70 characters.
- Title content should be related to its page content.
- Avoid duplication of keywords in title tag in term to get ideal keyword density.
- If you want to write multiple keywords in Title, use "|" sign to separate them.
- Title tag should be unique for every page related to its content.

Meta Tags

- Description tag must include a concise keyword oriented statement consists of two or more sentence about your website with range between 155-200 characters maximum.
- Avoid duplication of keywords in description tag in term to get ideal keyword density.
- Description tag should be unique for every page related to its content.
- Don't copy the content from the page to put in description tag.
- Use your main 4-7 targeted keywords in keywords tag.
- Use Meta Robot tags to control your website indexing and caching.

ALT & Heading tags

- If you are using any image in the website, use keywords in the Alt tags. Search engine cannot read or cache images but the code/tags behind the images.
- Do not use filenames like "image.jpg", "img.gif", and "a.jpg".
- While using links as image, always add alt tag and store images in their related directory.
- Use heading tags in place of headings in the content by using h1 to h4 maximum level.
- Does not use excessively heading tags in a page which can be considered as spam content.
- Avoid adding meaningless text in the headings tags who also not related to the structure of the page.

Website Structure

- While using most of the CMS applications in web development, URLs contain session ids and symbols like "?"," id=","&" etc which are not good in term of search engine indexing. To avoid this web site developers re-write URL by accessing .htaccess file to make then proper and search engine friendly. Many open source tools and websites available to help developers in writing .htaccess file.

- Avoid duplicate URLs on the same page and generic names like page.php.
- Create a directory folder for each content group/category, for example, an article website which can have multiple categories. For example, Website has category of clothes then the web pages related to clothes should be placed in Clothes category. The ideal webpage will be like that
  http://www.yourwebsite.com/clothes/shirts.html

- Ideal Directory/URL structure is two or maximum three level maximum. That's mean "/" should be two or three maximum in a URL.
- Adding Breadcrumb trail on every page makes navigation system better, search engine friendly and user friendly.
- Develop your website pages (particularly deep pages) with one click away from home page. Home page of the website is the most important page. This will make crawling process of deep pages is more easy.
- Develop a sitemap of the website, and update when add a new page. Submitting sitemap to the Google webmaster tool and other search engines webmaster tools help to get all webPages addresses to crawl and index but them.

Content Optimization

Content is very important part of any website. Since Google and other search engines store content of the website to display search results. With the latest update of Google search engine which named Google panda content is the most important part in search engine optimization process. Following is the key things while using content on the website

- Write interesting content to engage the reader. The major search engines are focusing on producing excellent possible user experience. That's means sites that present related information stand to achieve most visibility.
- Write the content which is relevant, interesting and shares that across social media.
- Keyword density is very important factor, so keep in mind it should be around 5% per page.
- Apply keyword tweaking features in content.
- Avoid adding too much keywords and irrelevant content to the title and description in a page. Google considers its spam and hence will give no value.

8.3   Off-Page SEO Outcomes & Recommendation

Link Popularity

Building back links of the website is very essential to have good ranking, PageRank as well as getting web traffic which depends upon search engine ranking. Three key things are important in a link building campaigns which are

1. Which websites and topic you are targeting
2. PageRank of the back links websites
3. Budget of the link building campaign

Before starting the link popularity campaign plan and make a link building strategy in which you decide which kind of link building you will do. For example, one way or three way linking will be beneficial while starting a campaign from scratch. After your website ready, first thing to do is to submit it to major search engines like Google, Yahoo and MSN by exploring their "add URL" pages. There are also many of websites on internet available for search engine submission. You can find the by writing "search engine submission" in Google search box. Following are the key things to focus in a link building campaign

- Always check the back link page weather it allows "do follow" link, is not a spam page and have maximum 20-25 back links. Check when the page is indexed by writing query cache:webpageaddress in Google search bar. Pages which are not indexed regularly are useless for link building.
- Link building process should be manual and should not be automated.

- It is recommended that PageRank of the back links should be 2+. Having high PageRank back links will create great chances to get rank quickly your website and vice versa.
- Relevancy of the content is very important. For example getting links for a real estate website, websites related to real estate, property or mortgage will be ideal in the link building campaign.
- All back links must have different IP address, preferably c class IP's is most beneficial.
- Back links pages should not use java scripts, framed pages, redirects, cloaking ,jump scripts, dynamic links, flash sites/pages, hidden links, no string parameters or database driven pages , CSS redirects, 301 redirects.
- Always add your target keywords in back link Title and in description. For example, you want to get rank on real estate California; then your back link title should be "Real estate California".
- Avoid getting back links in bulk at a time. Getting back links in huge amount can cause Google block or panelized.
- Buying paid links or link baits from home page give huge worth in link building campaign because home page of website crawls by search engines quickly. On the other hand it is expensive but cost effective.

While adding comment on forums, social media networks and blogs check do follow and indexing of the page where you are sharing content. Submitting press releases to good PageRank website are effective in getting more visitors and page indexing.

Note: If you want your important page to index quickly then submit it to top social media websites as well as if you can get home page back link will be most beneficial and vice versa.

If you are optimizing a blog or forum, use RSS submissions whenever you update or add new page or news. It will help for search engines to find your updated pages quickly.

8.3.1.1 SEO Tools and Maintenance

There are numerous SEO tools available on internet. Most of them free of cost or some good payable tools like internet business promoter, SEO Elite, digital point. After building your website the first thing is to verify it with w3 validator on validator.w3.org. Google provides two excellent tools i-e webmaster and analytics for website maintenance. Submit the website to Google webmaster to verify and submit sitemap in xml format. Using Google webmaster tool you can see what pages of your website are indexed by Google, back links crawled by Google with date, queries data by users to reach your website. Google analytics is used to analyze web traffic on your website. By using this tool web traffic information and data is collected with any features like location, page views, bounce rate, search engine sources, queries and keywords used to reach your website, conversion rate, bounce rate, internet browsers on daily, weekly and monthly basis. URLs of popular tools and softwares on Internet are following

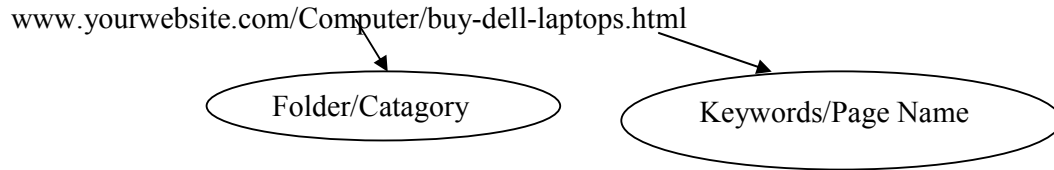| SEO Tools WebSite	Description/Use
www.google.com/webmasters	Google webmaster
www.google.com/analytics	Google Analytics
https://adwords.google.com/select/KeywordToolExternal	Keyword Research Tool
www.google.com/doubleclick/	Web Traffic indicator tool
http://tools.seobook.com/	SEO Tools
www.digitalpoint.com	Paid SEO Tool
www.mattcutts.com/blog	SEO latest Updates by matt cutts
www.highrankings.com	SEO latest updates by Jill Whalen
www.seochat.com	Latest SEO updates
www.seomoz.org	SEO updates and Tools

Table 9 SEO tools and important websites

8.4 SEO for Content management Systems

Evolution of content management systems has played a great part in website development. There are many open source like Joomla, WordPress, Drupal and paid enterprise CMS like Agility, Axcms.NET. Most of CMS systems are not developed according to SEO standards. There is always a need to add and develop SEO features in current CMS applications available. Before building a website with CMS integration it is necessary to analyze the requirements and SEO needs for the online web you are starting. These requirements are which CMS you will choose for development for which kind of application e-g blog or shopping cart or community website.

Most of the CMS solutions provide various SEO plugins and add-ons to develop search engine friendly. Choosing CMS is a critical process while building website for your online business. The check the quality of CMS before analyze its features is to validate by using w3.org validate utility. A CMS should provide the following key SEO features in development process

- Title and Meta Tag customization: A search engine friendly CMS should offer title and Meta tags to be customized on page-specific (i-e for every unique page) and particular sections of website such as website directories/folders pages. It means every web page title and Meta tags can be given separately to insure uniqueness of content of the page.
- It should offers to add and edit extra functionality on HTML tags such as "nofollow", robot tags on links and heading tags.
- Since URLs are the key in website structure. A good CMS must offer custom URL formation. The resultant URL should be static and SE friendly (i-e must not contain

string IDs, database driven,:, ? symbols ,_, etc). For example, a good URL can be following

www.yourwebsite.com/Computer/buy-dell-laptops.html

Folder/Catagory

Keywords/Page Name

An example of search engine friendly CMS in which web page editor in figure
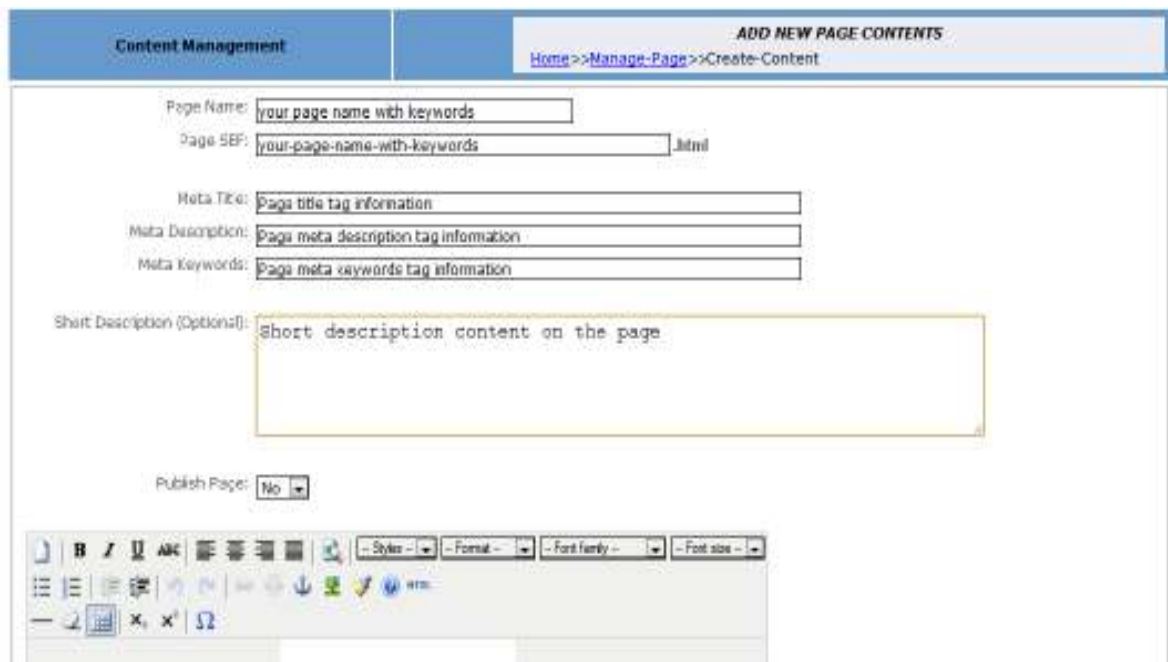

Figure 13 (CMS web page editor (source: own development project newsmediamatter.com))

- CMS should provide anchor text flexibility and customize navigation panels. A CMS should avoid pagination[4] which causes duplication of the content. Also a lot of CMS systems do not offer proper re-direction of expired and duplicate content. 301 re-direct is used for permanent redirection of the page.
- Control of alt images and CSS manual exceptions while editing and modifying elements of structure of the website and images code.

---

[4] Pagination is the process of dividing information (content) into discrete pages, either electronic pages or printed pages.

8.5   Comparison of Drupal, WordPress and Joomla SEO Features

8.5.1   Drupal

Drupal is amazingly powerful and easy development tool offers a variety of various options to make website search engine friendly. The Drupal offers many SEO plugins and functions to ensure maximum visibility of your website in major search engines. Some popular SEO modules are

Pathauto: This module automatically produces path aliases for different of content without involving the user to manually give the path alias.

Nodewords: It allows adding Meta tag pages.

Path Redirect: it uses for re-direction of URLs.

Content construction kit (CCK): It uses to create advanced custom views and new content types.

8.5.1.1 Pros and Cons of Drupal SEO

Following are the Pros of using Drupal CMS

- Drupal is open source, easy to install and reliable on all kind of webserver.
- It offers CSS enable web themes free of cost.
- Drupal provides a huge range of community developing modules and technical help.
- Drupal cache settings are easy and fast which boost caching process and hence good for the website ranking.
- Drupal is best option to develop community websites and social media networks applications.
- Drupal provides clean URL structure automatically hence without accessing rights of .htaccess file you get SE friendly URL structure.

Cons of using Drupal CMS

- Drupal is a heavy CMS in term of memory and processing power. For SEO friendly website page loading time and size of page matters. Drupal in somewhat has disadvantage regarding this feature.
- While adding and customizing advance features on website code, the code tends to become complex and hence it is difficult to modify code for further SEO enhancements of deep pages.

8.5.2 Joomla

Joomla is a multifunctional content management system and designed for large and complex websites such as shopping carts. Joomla development is a bit complex but exceptionally flexible. It is easy to create templates by using Joomla. Before starting your website development Joomla offers you set SEO settings. Joomla provides search engine friendly URLs however for deep pages and categories you need to re-write .htaccess file. Following figure 15 shows SEO setting panels in Joomla
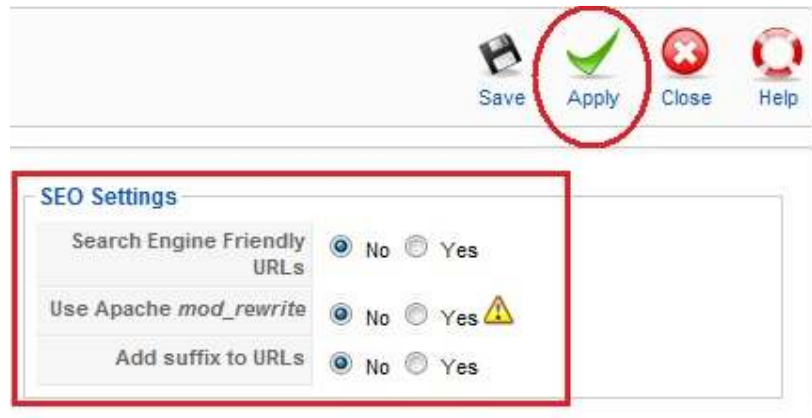

Figure 14 (SEO Setting Panel in Joomla (source: joomla.org))

8.5.2.1 Pros and Cons of Joomla

Pros

- Joomla provides SEO friendly templates and secure development procedures.
- You can set manually SEO settings and can control URL structure according to your requirements.
- It has a huge community which updates daily latest plugins regarding SEO and development.
- Joomla is beneficial for large application and massive database driven websites where you need to index millions of pages.

Cons

- It is difficult to understand its features and complexity; hence experienced developers need to implement changes such as SEO changes.
- By time version updates has caused a problem for old version users in modifying and adding plugins.

8.5.3 WordPress

WordPress is the leading CMS in term of usage among all others. WordPress is basically designed for blogging and no doubt currently the most popular CMS on the internet. It offers easy installation process, search engine friendly pages, utility to use .htaccess to produce static URLs which called permalinks, blogrolling and pinging. WordPress by default offers unique title and Meta tags modification for each page. It provides fresh content and best directory structure which is very helpful in search engine crawling. It offers commenting on posts which keep update the content on the page. WordPress provides a great number of themes and plugin than any other CMS from them most of are free and easy to implement. On WordPress platform everyday new applications and updates are announced especially regarding SEO. Using WordPress build web pages lighter with CSS extension which eventually rank higher in search engines.

8.5.3.1 Pros and Cons of WordPress SEO

Pros:

- WordPress is best for blogging which offers huge amount of plugins.
- Easy to install and crawl quickly.
- As content is the king in SEO, WordPress provides the best representation and control of content.
- WordPress' internal logic and link structure make search engine crawling fast.
- It provides excellent use of heading and image tags.
- It is very reliable to construct RSS feeds and pinging updates.

Cons:

- WordPress is not made for large applications such as shopping carts where huge amount of database queries conduct on daily basis.
- While accessing and updating .htaccess sometime WordPress creates problems for the developers to build search engine friendly URLs.

8.6 Advantages of Search Engine Optimization

Following are the main advantages of SEO:

- The primary advantage of SEO is that it is cost effective and gives long term ROI. When you ranked in top 10 on your desired keyword then you get maximum chance that use click on your website.
- SEO improve your website visibility to create a brand of your website.
- Return on investment of SEO is very much higher than the paid advertising.
- By using SEO tools you can manage the user traffic and track your leads for future decision making process.

- SEO make yourself or your organization social, active and reliable among internet communities. For example, in mobile industry GSMArena is become the key website for information and latest news.
- The coming era is of online business as it is growing rapidly. Therefore, need of SEO is essential for every online business.
- SEO makes your website globally accessible and hence you able to get users or customers from multicultural and multi languages societies.
- SEO offers lower administrative cost and generate much larger financial savings.

8.7 Disadvantaged of Search Engine Optimization

- Search engine optimization is a long term process which needs patience, time, money and right implementation of SEO tools and techniques.
- Since Google and other major search engines rapidly change their ranking algorithms. That' why there is a great risk involve in short term or poor SEO campaigns.
- SEO can increase web hosting cost. For example, to control huge web traffic you need extra bandwidth in your web hosting package.
- There is no standard among SEO, however Google and other major search engines has defined their rules and regulation which differ to each others. It is a great need for search engine practitioners to sit and define a unique set of standards of SEO.
- SEO requires continuous update of data and information on your website.
- SEO is not very friendly for graphics, image and flash base websites.
- SEO can generate too much leads/sales which can involve unwanted leads or unwanted users.

9 CONCLUSION & FUTURE WORK

In reference to above cited chapters, search engine optimization methods and tools are explained briefly and how it can be practiced on CMS applications. The overview given about search engine indexing and working tells how search engine works in background and how it gets the results while answering user queries. This thesis can be used as a primary guideline for any kind of SEO campaign. If you follow step by step the SEO methods and techniques as explained in this thesis, then you have a great chance to get ranked on major search engines.

Search engine optimization field is growing rapidly with respect to growth of online business. Search engines update their ranking algorithms frequently with the passage of time however major search engines like Google, Yahoo and Bing have different ranking criteria. White-hat SEO is practiced by most of SEO practitioners rather Black-Hat SEO which is rarely practiced. Keyword research and on-page optimization are the key thing in any SEO campaign. If on-page optimization is not implemented correctly on the website, off-page optimization and desired ranking results cannot be achieved. To measure the performance of SEO activities and website SEO maintenance, Google Webmaster and Analytics are the key tools to be used. SEO strategies are kept on changing with passage of time as search engines update their ranking algorithms. The latest SEO strategy is more about content and its representation rather than traditional SEO of putting keywords and getting links. Now those websites are ranked higher by Google who are more liked by users and have not only meaningful optimized but also interesting content.

Content management systems are involved in every kind of web system. Developing website with content management system can greatly boost organizational structure, branding and allows you to better optimize it for search engines. Most common CMS systems used for web development are Drupal, WordPress and Joomla. WordPress is most widely CMS among all other popular CMS systems and generally use for blogging and personnel websites. Implementing SEO on WordPress is easy among other mentioned and it offers many SEO plugin. The most popular plugin is "WordPress SEO". Drupal CMS is used of large applications particularly for community and complex websites. PathAuto is most powerful SEO extension for Drupal CMS. Joomla CMS is popular to develop huge database driven application such as shopping carts. AceSEF and sh404SEF are the most popular SEO plugins for Joomla. Implementing SEO in deep pages on Joomla website, for example a shopping cart website such URL like http://www.joomlawebsite.com/index.php?view=category&id=5,subcatagory?folder=&id= 3 will be tricky for developer to make it search engine friendly and one click away from home page.

Knowledge of search engine indexing and caching is very essential for every SEO practitioner and web developers. Search engine optimization depends upon indexing of web pages. If your key web pages are not indexed by major search engines they cannot get ranked. Search engines like pages which update on regular basis, having unique optimized content and quality back links.

Following are possible direction for future work in internet marketing and CMS applications area.

- Major search engines like Google, Yahoo and Bing have their own unique criteria of ranking. For example, Google prefer content creditability and user interest, Yahoo gives too much worth to back links particularly back links from its directory and partner websites. Search engine developers should research to design and development better algorithms of ranking results which are more specific to the user needs. The algorithms should be distinct because users have different interests.

- SEO tools should be available on one place which is compatible and accessible for all kind of search engines.

- Search engines should also work to rank better websites which use high graphic, development pattern MVC and AJAX based websites and reduce the delay time while providing SERPs.

- Major outsourcing CMS groups like Drupal, Joomla, WordPress and Mambo and other paid like Axcms.net, Agility etc should work on directory/URL structure and basic SEOs principal so that every client gets a optimized website after installation which save additional SEO costs. In future, those CMS will be successful who are ranked well with better SEO implementation and will have more web traffic.

10 REFERENCES

Adam, R. I. (2010). E-Marketing Strategy for Businesses. IEEE International Conference on E-Business Engineering (p. 434). Conventry: Department of Computing and the Digital Environment, Faculty of Engineering and Computing,Conventry University, UK.

Amanda, B. J. (2004). Web search: public searching on the Web (Vol. 6). (V. 6. Management, Ed.) Springer.

Ao-Jan, Y. C.-K. (2010). How to Improve Your Google Ranking:Myths and Reality. 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (p. 57). Evanston: Northwestern University, Evanston, IL, USA.

Bidgoli, H. (2010). MIS 2010. Cengage Learning, 2010.

Blankson, S. (2008). Search Engine Optimization (Seo): How to Optimize Your Web Site for Internet Search Engines : Google, Yahoo!, Msn Live, Aol, Ask, Altavista, Fast, Gigablast, Snap, Looksmart, and More (illustrated ed.). USA: Lulu.com, 2008.

Boiko, B. (2005). Content Management Bible (2, illustrated ed.). the University of Michigan, USA: Wiley Pub.

Brazell, A. (2010). WordPress Bible. John Wiley and Sons.

Cao, W. Y. (2010). Using Content Management System Joomla! to build a website for research institute needs. Management and Service Science (MASS), 2010 International Conference on (p. 3). Wuhan: University of China.

Chris, J. W. (2001). PageRank Explained or "Everything you've always wanted to know about PageRank". USA: searchenginesystems.net.

Chung, B. K.-S. (2001). A Method for Indexing Web Pages Using Web Bots. Rensselaer Polytechnic Institute, Troy, N.Y. 12180-3590, USA (p. 6). New York : IEEE computer society, pp1-6.

Crowder, R. S. (2009). Using Joomla: Building Powerful and Efficient Web Sites. New York: O'Reilly Media, Inc.

Danny, E. D. (2011). Search Engine Optimization (SEO) Secrets (illustrated, annotated ed.). John Wiley and Sons, 2011.

Dave, F. E.-C. (2009). Internet marketing: strategy, implementation and practice (4, illustrated, annotated ed.). USA: Pearson Education, 2009.

David, V. (2008). Get to the Top on Google: Tips and Techniques to Get Your Site to the Top of the Search Engine Rankings (illustrated ed.). Nicholas Brealey Publishing, 2008.

Drupal. (2011, 05 15). pathauto. Retrieved 05 25, 2011, from Drupal.org: http://drupal.org/project/pathauto

Drupal, O. (2005, 05 05). Drupal about us. Retrieved 05 19, 2011, from Drupal.org: http://drupal.org/about

Drupal, W. (2011, 05 21). 10 Reasons to use Drupal CMS. Retrieved 05 21, 2011, from Webmaster tips, Drupal tutorial: http://tips.webdesign10.com/drupal/about-drupal-311.html

Enaceanu, A. (2007). SEO Techniques for Business Websites. Informatica Economica nr. 2 , 77.

Enge, S. S. (2009). The Art of SEO (illustrated ed.). O'Reilly Media, Inc., 2009.

Eric, S. S. (2009). The Art of SEO (illustrated ed.). O'Reilly Media, Inc., 2009.

Evan Bailyn, B. B. (2011). Outsmarting Google: SEO Secrets to Winning New Business. USA: Que Publishing, 2011.

Farida Hasanali, P. L. (2003). Content Management: A Guide for Your Journey to Knowledge Management Best Practices. APQC, 2003.

Fleischner, M. H. (2011). Seo Made Simple: Strategies for Dominating the World's Largest Search Engine (2 ed.). SEO Made Simple, 2011.

Forgan, C. N. (2009). Drupal and SEO. (S. LLC, Ed.) Retrieved 05 30, 2011, from Silverfoxinteractive: www.silverfoxinteractive.com/pdf/drupal-ppt/Drupal-and-SEO.ppt

Gautam, P. S. (2004). Crawling the Web. Web Dynamics: Adapting to Change in Content, Size, Topology and Use (p. 178). Berlin: Mark Levene, Alexandra Poulovassilis (Ed.).

Gifford, J. (2010). Digital Public Relations:E-Marketing's Big Secret. (continuing higher education review, Ed.) E-Marketing's Big Secret , 74, 72.

Google. (2011, 03 20). URL structure. Retrieved 05 2011, 10, from Google webmaster tools help: http://www.google.com/support/webmasters/bin/answer.py?answer=76329

Graham, B. K. (2008). Key Differences between Web1.0 and Web2.0. First Monday, Volume 13 Number 6 - 2 June 2008 , Volume 13 , 1-30.

Hayman, D. R. (2007). How to Maximize Traffic to Your Website. USA: Expert Software Systems, 2007.

Hui, C.-h. L. (2010). Optimizing the Web Search Engines With Features and Caching. International Conference on Web Information Systems and Mining (pp. pp.193-197). Sanya: IEEE.

II, T. D. (2008). Advanced Search Engine Optimization: A Logical Approach. Lulu.com, 2008.

Interactive, L. (2010). 2010 CMS Intelligence Report. Dallas: LevelTen Interactive, 2010.

James, K. (2011). Life outside the Google top ten. Retrieved 02 05, 2011, from Kilroy James: http://www.kilroyjames.co.uk/2008/05/life-outside-the-google-top-ten/

Jerkovic, J. I. (2009). SEO Warrior. O'Reilly Media, Inc.

Jerkovic, J. (2009). SEO Warrior (illustrated ed.). O'Reilly Media, Inc., 2009.

Jones, R. (2011). Keyword Intelligence: Keyword Research for Search, Social, and Beyond. Indiana: John Wiley & Sons.

Keil, A. (2011). How to get noticed in Google: Introduction to Search Engine Optimisation in the Era of e-Retailing. GRIN Verlag, 2011.

Kernek, A. (2005). Put Your Business Online. Pacifica Endeavors LLC, 2005.

Kothari, D. C. (2008). Research methodology: methods and techniques. USA: New Age International, 2008.

Kristopher, J. (2010). Search Engine Optimization: Your Visual Blueprint for Effective Internet Marketing (2 ed.). John Wiley and Sons,.

Leary, S. (2010). Beginning WordPress 3. USA: Apress, 2010.

Ledford, M. E. (2009). Google Analytics (3, illustrated ed.). USA: John Wiley and Sons, 2009.

Lopuck, L. (2006). Web design for dummies (2, illustrated ed.). For Dummies.

Malaga, R. A. (2008). Worst practices in search engine optimization. (N. J. Montclair State University in Montclair, Ed.) Worst practices in search engine optimizatio , 51, 150.

Michie, J. (2007). Street smart internet marketing: tips, tools, tactics & techniques to market your product, service, business or ideas online. USA: Street Smart Internet Market,2007.

Morley, D. (2009). Understanding Computers: Today and Tomorrow, Comprehensive. Cengage Learning, 2009.

Musser, J. (2006). Web 2.0 Principles and Best Practices. O ' R e i l l y R a d a r Inc. .

Neglia, L. F. (2009, 01 01). Drupal and SEO. (S. LLC, Ed.) Retrieved 05 30, 2011, from Silverfoxinteractive.com: www.silverfoxinteractive.com/pdf/drupal-ppt/Drupal-and-SEO.ppt

Odom, S. (2011). SEO For 2011: Search Engine Optimization Secrets. USA: MediaWorksPublising.com.

organization, S. (2008, 02 26). Sitemaps.org. Retrieved 05 08, 2011, from Sitemaps.org: http://sitemaps.org/protocol.php

Page, S. B. (2001). The Anatomy of a Large-Scale Hypertextual web search engines. Stanford University, Computer Science Department. Stanford: Stanford University.

Pokorny, J. (2004). Web searching and information retrieval. Computing in Science & Engineering , 4, 48.

portfolios, E. (2008). web 2.0 class. Retrieved 05 12, 2011, from electronic portfolios: http://electronicportfolios.com/web2/class/index.html

Potts, K. (2007). Web Design and Marketing Solutions for Business Websites. Apress.

Randy, J. B. (2006). Embedding Search Engine Optimization tactics during CMS deployment. Ottawa: non-linear creations.

Rice, S. L. (2009). Mastering Web 2.0: Transform Your Business Using Key Website and Social Media Tools (illustrated ed.). Kogan Page Publishers, 2009.

Ryals, K. (2008). Seo 100 Success Secrets - Search Engine Optimization 100 Most Asked Questions: The Missing Seo Tools Handbook and Guide to Ranking Factors. Lulu.com.

Saini, R. K. (2011). A Study on SEO Monitoring System Based on Corporate Website Development. International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), , 1 ,No.2, 49.

Sarina, S. S. (2008). Web caching and prefetching: What, why, and how? Information Technology, 2008. ITSim 2008. International Symposium on , 8.

Savan, V. J. (2011). Performance Analysis of Content Management Systems- Joomla, Drupal and WordPress. (I. Journel, Ed.) International Journal of Computer Applications (0975 – 8887) , 21-04.

Seo, c. (2011, 05 05). robot meta tags. Retrieved 05 27, 2011, from www.seoconsultants.com: http://www.seoconsultants.com/meta-tags/robots/

Silverman, D. (2011). IAB Internet Advertising Report. IAB, PWC www.PWC.com.

Small, M. (2005). SEO Answer Book. Lulu Enterprises Inc., 2005.

Stanek, W. R. (2001). Microsoft FrontPage 2002 unleashed (illustrated ed.). Sams Publishing, 2001.

Szymanski, M.-S. C. (2001). A Method for Indexing Web Pages Using Web Bots. In IEEE (Ed.), Proc. Int. Conference on Info-Tech & Info-Net ICII'2001, (p. 6). Beijing, China: Department of Computer Science, Rensselaer Polytechnic Institute, Troy, N.Y.

Thorson, K. (2004). Modeling the Web and the Computation of PageRank. Roanoke: Hollins University.

Tonkin, S. (2010). Performance Marketing with Google Analytics: Strategies and Techniques for Maximizing Online ROI. John Wiley and Sons, 2010.

Torbert, M. (2011, 10 12). WordPress-all in one SEO Pack. Retrieved 12 13, 2011, from http://wordpress.org: http://wordpress.org/extend/plugins/all-in-one-seo-pack/

Townsend, R. J. (2010). Foundation Drupal 7: Learn how to Use the Drupal Framework to Quickly Build Feature-rich Web Site. New York: Apress, 2010.

Viney, D. (2008). Get to the Top on Google: Tips and Techniques to Get Your Site to the Top of the Search Engine Rankings (illustrated ed.). Nicholas Brealey Publishing, 2008.

w3c. (2011). W3c website standards. Retrieved 03 07, 2011, from W3C: http://www.w3.org/standards/

W3c. (2011). W3c website standards. Retrieved 03 07, 2011, from W3C: http://www.w3.org/standards/

Walter, A. (2008). Building findable websites: web standards, SEO, and beyond (illustrated ed.). Peachpit Press.

Waring, J. M. (2010). The Official Joomla! Book (illustrated ed.). New york, USA: Addison-Wesley Professional.

WordPress. (2000, 08 25). SEO Ultimate. Retrieved 08 29, 2011, from Wordpress.org: http://wordpress.org/extend/plugins/seo-ultimate/

Wu Di, L. T. (2010). Study on SEO Monitoring System Based on Keywords & Links. (D. J. lSoftware Technology Institute, Ed.) Study on SEO monitoring system based on keywords & links , 5, 455.

Zhang, Y. (2005). Web technologies research and development. 7th Asia-Pacific Web Conference, APWeb 2005 (p. 1082). Shangai: Springer,.