Search Engine as a Tool for Destination Choice: A Paradigm shift

Sanjeev Dhawan¹, Rakesh Kumar², Divya Sethi³

¹Faculty of Computer Science & Engineering, University Institute of Engineering & Technology (U.I.E.T), Kurukshetra University, Kurukshetra (K.U.K)- 136 119, Haryana, INDIA.
²Faculty of Computer Science, Department of Computer Science and Applications (D.C.S.A), Kurukshetra University, Kurukshetra (K.U.K)- 136 119, Haryana, INDIA.
³GM Conferencing & VSAT Solutions, Enterprise Services, Bharti Airtel, Gurgaon, Haryana 122 001, INDIA.

Abstract: In this paper, we present the different scenarios of search engine optimization, a paradigm shift in the market of tourism science & management. Promoting websites online can acquire various forms. Nevertheless, the most popular ways to advertise the tourism based websites are through search engine optimisation pay per click, pay per impression, and pay per percentage advertising techniques. However, each of these methods has one aim; to catch the people searching for your product onto your website in an optimised way. In correct optimisation the chances of getting a website found under its key phrases should be significantly increase. If your website can’t be found in the web space priorities then potential customers will visit a competitor’s site and you could lose out on a faith and sale. A thoughtful SEO strategy requires more than an automated service promising submission to “10000” search engines available today. This paper addresses the function & design prediction of SEO techniques, which will be used to increase the web page ranking among the customers, presents the implementation of a practical web based information system, by using the PPC, PPI, & PPP design techniques, and discusses the implications & results of SEO techniques with the use of above mentioned approaches.

Keywords: Function and design prediction of SEO, implementation of web based system, implications of SEO techniques, search engine algorithms

1. Introduction

Conde Nast Traveller ranked India as the fourth most preferred travel destination & Lonely Planet selecting the country among the top five destinations from 167 countries this year. India has finally made its mark on the world travel map. Global recognition is evident from the rising number of India-bound tourists. From just 17,000 international arrivals in 1951, the number has grown to 39 lakh in 2005. According to the Ministry of Tourism, we have already clocked almost 30 lakh foreign tourists between January & September, as compared with 27 lakh during the same period last year. The tourist arrivals, pushing India's foreign-exchange earnings from $3.5 billion in 2003 to $5.7 billion in 2005. Much of the credit goes to the Ministry of Tourism's 'Incredible India' campaign launched towards the end of 2002 [Saikat Neogi 2006]. The campaign built a combined effort in international print, electronic, & most importantly in Internet media to showcase the country's tourism-friendly aspects. Earlier, the ministry was solely dependent on its overseas offices to promote India as a tourist destination. But now we have one of the top-most mediums for promoting the tourism i.e. through user-friendly Websites techniques, which have been discussed in this paper.

The volume of information on World Wide Web is increasing enormously everyday. The primary tool to locate any information on the web is search engine. The goal of search engine is to locate relevant documents in response to a user’s query. Documents are typically retrieved as a ranked list, where the ranking is based on estimation of relevance [Belkin, N., & Croft, W.B. 87]. There may be millions of the sites pertaining in some manner to user’s queries, search engine will retrieve them, rank them on the basis of their relevance using some criteria, & present a huge list of documents. But studies show that approximately 80% of Web searchers never view more than first 10 results in a result’s list [Holscher C. & Strube G. 2000, Silverstein C., Henzinger M., Marais H., & Moricz M. 1999, Jansen B.J., Spink A., & Saracevic T. 2000]. A search engine typically displays results page by page, perhaps 10 hits at a time. Many users tend to look at only the first page of results they get from a search engine, rather than scrolling through pages & pages of hits. Some users scan additional result pages, usually in the natural ordering. A search session implicitly terminates when the user decides not to browse additional result pages on the topic that initiated the session.

Several studies have analyzed the queries that users submit to search engines, & the length of search sessions [Markatos, E. P. 2000, Lempel, R., & Moran, S. 2003]. Three findings that these studies share are particularly
relevant to this work are: (i) The queries submitted to WWW search engines are very short, averaging less than 2.4 terms per query, with over half of the queries containing just one or two terms. These results were reported by [Croft, W. B., Cook, R., & Wilder, D. 1995] & [Wen, J.R., Nie, J.Y. & Zhang, H.J. 2002]. While the two studies define query terms somewhat differently, the reported term counts may be loosely interpreted as the number of words per query. (ii) Users browse through very few result pages. These studies differ in the reported distribution of page views, but agree that at least 58% of the users view only the first page (the top-10 results), & that no more than 12% of users browse through more than 3 result pages. While the above describes the behavior of users as they browse through multiple result pages, statistics have also been gathered on the browsing patterns of users as they view a single page of results. It has been observed that users are reluctant to scroll beyond the visible part of the page, & so search results that are “above the fold” are viewed (and clicked on) by more users than results at the bottom of the page [Broder, A. 2000], and (iii) The number of distinct information needs of users is very large, as can be seen from the huge variety of queries submitted to search engines.

However, popular queries are repeated many times, & the 25 most popular queries account for over 1% of all queries submitted to the engines. With the growing economic impact of the WWW, & the growth of e-commerce, it is crucial for businesses to have their sites ranked high by the major search engines. In the increasingly market-driven environment that is the web, it is becoming even more important to ensure that, as many people should see the site as possible. In particular, corporate & organizations who are trying to attract advertising revenue to their web sites need to ensure that they can show potential advertisers that placing an advertisement on their pages is likely to achieve maximum exposure. There are quite a few companies who sell this kind of expertise; they design web sites, which are tailored to rank high with specific queries on the major search engines. Under such circumstances, web authors are increasingly beginning to rely on underhand techniques to ensure their sites get seen. Emerging practices are popularly known as keyword spamming (Keyword spamming is doctoring the content of a web site to ensure that it hits the top of a list of results retrieved from a web search engine. It is writing keywords multiple times so that one will be on top of a search results) & site optimization. These companies (which call their business “search engine optimization/positioning”) research the ranking algorithms & heuristics of term-based search engines & know how many keywords to place (and where) in a web page so as to improve the page’s ranking (which directly impacts the page’s visibility) [Lempel R. & Moran S., Salsa 2001]. A less sophisticated techniques used by some site creators, is called keyword spamming [Chakrabarti, S., Dom, B., Gibson, D., Kumar, S.R., Raghavan, P., Rajagopalan, S., & Tomkins A. 1999]. Here the author repeats certain terms (some of which are remotely connected to their site’s context), in order to lure search engines into ranking them highly for many queries.

So it has become all-important to be ranked in top-10. To design user friendly & high-quality Websites require lot of efforts. So lot of concentration will be given to design a precise & reliable Website depending upon the Search Engine Optimisation (SEO), Pay per Click (PPC), Pay per Impression (PPI), & Pay Per Page (PPP) techniques.

II. Search Engines

It is a well known truth that maximum users use search engines to locate the information on the web. Three components of search engine are – catalogue engine, spider, and query processor. The Boolean search engine uses exact matching to match documents to a user query by finding documents that are 'relevant' in terms of matching the words in the query. The adjective "Boolean" refers to the use of Boolean operators AND, OR, & NOT to logically combine the query words. The Boolean information retrieval model considers which keywords are present or absent in a document or title. Thus, a document is judged as relevant or irrelevant -- there is no concept of a "partial match" between documents & queries. The inability to identify partial matches can lead to poor performance [Baeza-Yates, R., & B. Ribeiro-Neto 1999]. Other more advanced set theoretic techniques, such as the so-called "fuzzy sets", try to remedy this black-white Boolean logic by introducing shades of gray. For example, a title search for Tourism AND Hotels on a Boolean engine causes the virtual machine to return all documents that have both words in the title. As a result, a relevant document entitled “Tourism” will not be returned. Fuzzy Boolean engines use fuzzy logic to categorize this document as somewhat relevant & return it to the user. Fuzzy engines, on the pattern of human reasoning, use approximate information & uncertainty to generate decisions. The main limitation of Boolean search engines is observed in dealing with the problems of synonymy (multiple words have the same meaning) & polysemy (word has multiple meanings). Boolean search engines also require the user to be familiar with Boolean operators. E.g., to find information about the phrase Taj Mahal Fort, many engines require quotation marks around the phrase, which tell the search engine that the entire phrase should be searched as if it were just one keyword. A user who forgets this syntax requirement would be surprised to find retrieved documents about Taj Mahal tea or Taj Mahal mineral water for actual Taj Mahal fort. Nevertheless, Boolean model form the basis for many search engines. There are several reasons for their popularity: (i) Creating & programming a Boolean engine is straightforward. (ii) Queries can be processed quickly - a quick scan through the keyword files for the documents can be executed in parallel.
III. Search Engine Optimization vs. Pay Per Click

Promoting Websites online can take many forms, such as SEO, PPC, PPP, & PPI. Each of these methods has one aim: to catch the people searching for your product onto your Website. Although how they achieve this goal differs slightly.

A. Traditional SEO

SEO is the process of optimizing your Web site to reflect specific keywords & phrases that are relevant to your business & for which you want to attract visitors to your site who are searching for such words. This optimization relates to a variety of elements like title tags, meta tags, alt tags, link structure, link popularity & the content of the site itself, not only on your web site's home page, but its sub pages as well. Once your Web site is properly optimized, the goal then is to make sure that each of the top search engines find your site & add as many pages as possible in their indices. These engines will usually start with your home page & then work their way to other pages of the site. The important aspects of SEO are as follows: (a) Make sure that the sub pages or the main areas of your site are accessible from the home page. (b) It can be assumed that if many of your pages are optimized for different keywords & can be found in the search engines, they will draw traffic to your site. (c) SEO also involves making sure your site is listed in the major directories such as Yahoo & Open Directory to name a few. It is important to make sure you are listed in the right category & have the reflective keywords in title & description that user is trying to target. Other elements of SEO include monitoring your positioning in the search engines, making adjustments as necessary to your site to stay in top position & submitting to new engines that come along as well.

Figure 2.1 Framework flow diagram for SEO

Figure 2.2 Framework flow diagram for PPC

Figure 2.1 shows the framework flow diagram for SEO in which one should identify competition to select keywords that you expect surfers to use to locate your site, submit your obvious keyword terms & use the online freeware research tools to locate top ranked competing sites. The purpose of link tracking allows you to identify & choose possible sites that you will want as link partners once your site is content-rich & optimized. If these sites link to your competition, then you want them to also link to you. This is also done so that these potential link partners can be checked for meta keywords that might be related to your site that were missed by your competition. The purpose to identify keywords is to understand the search terms & keywords used to locate your site as a part of a search engine query. To help in this keyword analysis effort we offer advice to research the competition & query history information available through several sources. Repeat steps 1 through 3 until you have finished all your keyword searches. The purpose of combine keywords is based upon frequency of use & applicability. It is also important that you add words to help clarify the use of your keywords. For instance "Bank" might have very different meanings -- such as River Bank, Blood Bank, State Bank, Reserve Bank, etc. To add your keywords to the content is very much important to discover your real issue. Also, most sites use images to relay their message to humans, but are now discovering that search engines are blind & do not see images. The aim to tune keyword list will allow you to submit (to replace or add a page to the index) a URL once every day & changes are usually live quickly (faster than other search engines), so use the search engine with the fastest indexing process to 'debug' your keywords & SEO tactics. Once the same words are in your meta title, description & keywords, & also in other text & displayed content, & you are satisfied with your results, & only then, is it appropriate to submit your URLs to each major search engine. Furthermore, the Web site submission to search engine is done by making registration at search engine server & moreover time-by-time Web site checking has to be done. Finally, go to step 1 again & again for better search optimized results.
B. Pay per click search engine advertising

PPC is a service in which an advertiser selects specific keywords or phrases & then creates a listing that will show up when someone searches for that phrase. The advertiser selects an amount they are willing to pay for each click on their listing, which results in a visit to their site. They are the sponsored links you might see when you perform a search in a search engine. They usually appear at the top, bottom or right side of the search engine results page & are identified as sponsored links. At Overture, you can bid anywhere from $0.10 up to $50.00 for each click. If other advertisers have selected the same keyword or phrase as yours, you are competing against them for the highest position. Whosoever is interested to spend the most shows up first & the others following in order. What makes PPC attractive in the case of Overture is that they distribute their paid listings to other partner sites. In fact, if you bid in one of the top three positions at Overture, your listings will also show up at some of the leading search engines including AltaVista, HotBot, InfoSpace, iWon, Lycos, MSN, Yahoo & others. They also show up in several meta crawlers & other minor search engines. There are many other PPC programs including Google AdWords, Altavista, ah-ha.com, Sprinks & Looksmart to name a few. All function in a similar manner.

Figure 2.2 shows the framework flow diagram for PPC in which a discovery phase searches a word or phrase that people would employ to locate information on the products or services or topic that they are interested learning more about or purchasing. The core idea behind discovery phase is to identify the words or phrases that are important to your target audience & therefore will be the most effective to use on your Web site. The purpose of the design phase is to design the PPC Campaign based on the strategy & short or long-term needs identified in the discovery phase & help develop your bidding strategy. This phase could include: modeling the keywords in terms of monthly counts, estimated click-through rates, & conversion rates, identifying which PPC engines will be used for the campaign, identifying site landing pages & destination pages for conversion purposes, & identifying any tracking & reporting tools. The purpose of the deployment phase is to develop the PPC strategies & campaign design produced in the first two phases. This phase includes converting the audience issues into “creative content statements.” Once the creative content statements are finalized, the ads are customized for each PPC engine to adjust for each engine’s constraints, such as the length of the ad, the reach & demographics of the engine, etc., & bidding strategies are developed. At the conclusion of the deployment phase, you should: development of title & description by PPC Engine for each keyword, application of tracking URLs & reporting tools, development of new destination pages or editing existing pages, & implemented tracking & reporting tools. Finally, the purpose of the deployment phase is to “launch” & monitor the PPC Campaign as a result of the short & long-term needs. The deployment phase begins by uploading the PPC advertisements to the PPC search engines. Some PPC Engines allow advertisements to “go live” almost immediately & others require editorial approval. Finally, go to step 1 again for making sufficient changes to Web site depending upon the Web server statistics in term of their log files. The biggest benefit of PPC is the fact that it will provide you with an immediate boost of qualified visitors, lead & sales giving you fast results within just hours or days. In fact, a PPC advertising program is your best option if you seek fast results & a good return on investment while you are waiting for your SEO program to “ramp up”. PPC search engines also give you the added benefits of being able to quickly test your web site & track your alteration rates (leads, opt-ins, & sales) & turning keywords (visitors) on & off easily. So, as a short-term strategy PPC gives you the clear advantage over SEO. But, the disadvantage is the cost involved. Depending on the market demand for your keywords & clicks, your PPCSE campaign can generate tons of traffic & can cost hundreds, even thousands per day. With various optimization strategies you can control your costs, but over the long term natural search engine optimization will give you a higher return on your marketing dollar. The screen shot below shows the SEO, & PPC areas from Google using the example phrase ‘India Tourism Perspective’.

![Figure 4: screen snapshot of the SEO & PPC areas from Google.](image-url)
C. Google pay per click advertising

Google is the best-known PPC provider. With Google the PPC adverts, labeled 'sponsored links', appear on the right of the page, & sometimes also at the very top. You only pay when someone clicks on your advert. The minimum cost per click is five pence. Most of the time you will find that others also want their advert to show up for your keywords. That's where the bidding comes in. The more you are prepared to pay per click, the higher in the list of adverts you will appear. You can sign up to Google Adwords & within an hour you can start getting traffic to your Website. In designing a Website be specific. Don't use vague or general terms. The more precise & targeted your keywords, the less you will pay per click & the better quality of visitor you will welcome to your Website. Getting lots of traffic is not the objective. You only want traffic that is likely to turn into sales. If your company sells to the local area, make sure to include your town or county in the keywords. There is no point getting lots of people in New Delhi to your Website, if you only deliver to Chandigarh. For feedback mechanism for your Website, run two adverts & compare their results. The same can be done for e-commerce sites to measure, which adverts resulted in a sale by checking the log files from the respective Web servers.

D. Importance of SEO

With the current rise of pay PPC advertising services, professional SEO companies have to deal more & more with the issue of why their clients should spend money on traditional SEO when they can simply "buy their way to the top." Natural search engine optimization usually gives you a much higher return on investment than pay per click. This is true for two main reasons: (a) More searchers click the natural search engine results versus the PPC ads, so you'll get much more traffic for less, (b) One of the biggest factors to improve your rankings with natural search engine optimization is by boosting your "Link Popularity" by acquiring or paying for links that point to your Website. These links give you lasting results by giving you top rankings & traffic from the search engines. Plus, the links themselves will provide a significant boost in long-term traffic. It is also observed that more than 80% of all search engine traffic in July, 2006 came from the top 3 engines Google, Yahoo, & MSN. We have considerable reason to believe that your appearance in relevant search engines will boost your rankings in the 'Big 3' as well as gaining additional traffic from the smaller engines. The figure 3 shows the search engine relationship chart between top search engines, which has been designed virtually after carried out an empirical study with different people including students of the advances university classes, industry person, & facilitators.

Figure 3 shows the search engine relationship chart between top search engines

When a Website is implemented correctly, search engine optimisation can have a dramatic effect on a business. Many companies purchase a Website in hope that it will bring business without actively promoting it. Alternatively many companies are offering 'Web promotion' businesses for submitting your site to thousands of search engines in the hope it will make their site easier to find. In reality there are only half a dozen search engines worth spending time over, with Google, Yahoo & MSNbeing the biggest of these. We carried out an empirical study with the students of the advanced university classes that used various search engines & client-server based Web technologies for measuring & predicting the search behavior of the information retrieval systems. Our first goal was to compare the relative importance of each design activity. These students have Web log files on their computers, which monitor the sites they visit. These information are compiled to produce results. Figure 4 shows are statistics about the relative usage & importance of popular search engines. In order to assess the relative usage of popular search engines using the Internet services, it has been found (figure 4) that academic community use Google (68.3%), Yahoo (16.4%), Rediff (6.4%), Altavista (2.9%), Dogpile (1.6%), MSN (0.7%), AOL (0.6%), Excite (0.4%), Hotbot (0.3%), & remaining respondents use 2.4% of other search engines. Our first goal was to compare the relative usage & importance of major search engines. Second, we tried to assess the accuracy of the top search engines for optimized results in context to user queries. Third, we also discussed the implications & results of SEO techniques by capturing the dynamics involved in user behavior. The results obtained from the user assessment can help us to systematically identify the designing of tourism based Websites.
for evaluating priority levels, reliability, compatibility, usability, maintainability, complexity, cost, configuration, time requirements, types of interfaces, & type of nature of Website design.

Figure 4 shows the statistics of relative usage of popular search engines

E. Ethical SEO

Suppose there are 300 hotels in New Delhi, each of which has a Website. Against a query "New Delhi hotels", all 300 Websites are equally relevant. But Google & other engines have been designed to display the 10 results at a time. But which of the 300 hotel sites will be displayed in the first 10, in the second 10, & so on. It is well known that the sites that are nearer the top will take all the business, & those that are further down will get none. It is patently obvious that all 300 equally relevant Websites will not be displayed on the first results page. It is also obvious that equally relevant sites cannot be displayed where they belong. The site is just as relevant as the top ones; it’s just that Google cannot satisfy all the relevant sites. So there is nothing wrong by the owner of one of the Websites to try & push their sites to the top. This is where ethical SEO comes. SEO allows the pages of relevant Websites to be displayed at or near the top of relevant search results. SEOs have exactly the same aim as the engines – ‘relevant search results’. The difference is that search engines don’t care about individual Websites, whereas search engine optimizers & Website owners do. Engines don’t care if a particular Website is in the top 10; SEOs care very much that a particular Website is in the top 10. But they can’t get an off-topic site there because the search engine algorithms see to that. & that’s an important point - search engine optimization can only get pages to the top of relevant results. The search engines own algorithms keep off-topic pages out. As a search engine optimizer, we give Search Engines a choice & adjust things according to our Website design so that our relevant pages are at the top of the relevant results & that’s why PPC, PPI, & PPP come into the existence.

F. PPC, PPI, & PPP

Pay per click (or PPC), pay per impression (or PPI), & pay per percentage (PPP) are three distinct approaches to online advertising. PPC tools bring targeted traffic to a web site quickly & efficiently, whereas PPI tools can help to build brand recognition & PPP tools can help to build company’s revenue. The PPC charge advertisers for web site visitors; the PPI charge advertisers for message displays; & PPP charge advertisers for percentage of genuine message displays. The following table 1.1 summarizes the key differences between the three advertising approaches:

<table>
<thead>
<tr>
<th>PPC</th>
<th>PPI</th>
<th>PPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Objective</td>
<td>Build Web Site Traffic</td>
<td>Build Brand Name</td>
</tr>
<tr>
<td>Media Exposure</td>
<td>Search Engine Results</td>
<td>Web Page Banners</td>
</tr>
<tr>
<td>Cost per visitor</td>
<td>Low to medium</td>
<td>High</td>
</tr>
<tr>
<td>Cost per impression</td>
<td>Medium to high</td>
<td>Low</td>
</tr>
<tr>
<td>Pros</td>
<td>Predictable cost per visitor</td>
<td>Predictable cost per impression</td>
</tr>
<tr>
<td>Cons</td>
<td>Time-consuming bid management</td>
<td>Costly for targeted products &amp; services</td>
</tr>
</tbody>
</table>

Table 1.1 summarizes the key differences between the three advertising approaches

PPC advertising is, of course, one of the alternatives in the increasingly popular 'pay per action' marketing field. As visitor tracking systems become more sophisticated, two other online advertising models have been developed: 'pay per lead' & 'pay per sale'. These models are called affiliate marketing, can offer advertisers valuable exposure on a range of relevant websites & offer website publishers an opportunity to generate revenue.
from their resource sites. PPI & PPP are the examples of online advertisements & images or impression, which generally come while we browse our e-mail accounts (like Rediff, Yahoo, Gmail, hotmail etc.).

G. Click & impression frauds

One way to prevent click fraud is to revert to a PPI model. In this way, clicking on an ad does not defraud the advertiser. Click fraud is the biggest threat to the Internet economy. It has been shown that a simple system, PPP of impressions, is immune to both click fraud & impression fraud. In this system, an advertiser picks a keyword, e.g. “star hotel” & purchases, perhaps through bidding, a certain percentage of all impressions for that keyword. For instance, an advertiser might pay $1.00 to Google search. In return, the advertiser might receive 10% of all impressions for “star hotel” for 1 week. What does this mean? It means that for 1 week, one out of ten times that someone searches for the word “star hotel”, they will see the ad. If someone clicks on the ad, the advertiser is not charged any extra money. It might seem that this method is subject to impression fraud: someone could create fake searches for the word “star hotel” causing fake impressions. A common variation on the PPC model adjusts the pricing or the positioning based in part on click through rate. Some services, like Google, also have a minimum click through rate. This creates its own additional motivations for impression fraud. Since click through rate is roughly the ratio of number of clicks to the number of impressions. With weighted-pay-per-percentage, we first choose at random which word in a phrase to target, & then select it based on the percentage of volume purchased. For instance, if an advertiser purchases a weighted 10% share of the broad matches for “star hotel,” he will receive the following: 10% of exact match searches for “star hotel”, 5% of the searches for two word phrases that include star hotel, 3.33% of the searches for three word phrases that include star hotel, etc. So, in order to prevent fraud, we still must respect certain constraints. For instance, if we sell 70% of matches for “star hotel”, & 80% of matches for “seven star hotel”, then for a phrase like “seven star hotel”, we can sell no more than 25% of the traffic (100% - (70%/2 + 80%/2) = 25%). Notice that weighted-pay-per-percentage is immune to fraud, because it preserves all independence assumptions. Our preferred method simply always chooses the first word, or always chooses the last word. We’ll call this "Prefix or Suffix PPC”. In this system, rather than selling a full broad match, we sell only prefixes or suffixes. For instance, “star hotel *” would match “star hotel Delhi” & “star hotel Bombay” but not “wanted star hotel. For instance, “* star hotel Delhi” would match “seven star hotel Delhi”.

IV. Designing a high-quality Website

Designing a search engine friendly Web site is not complicated, but it requires dedication & constant communication between the graphic designer, content developer & SEO analyst, as well as special attention to the site’s architecture. Site architecture should not be an afterthought, nor should SEO. Your site must be both visually appealing to visitors & contextually appealing to the search engines. In order to attain both of these goals, we recommend enlisting a graphic designer to design the look & feel of the site, & a content developer to create a clear message for your site. This will cause the search engines to rank your site higher & will increase your visibility for visitors. Creating a site that is easily navigable, visually appealing, & gives your visitors what they are looking for, will encourage visitors to stay longer & come back often. We recommend that in at least 200 words as close to the top of your HTML code as possible, you need to make sure that you are using keywords in a manner that makes sense to a visitor & will allow the search engines to see your site as a subject matter expert. Never stuff your pages with keywords; the engines will penalize you.

A. Search engine algorithms

Each search engine employs different algorithms for determining page rank & these algorithms are subject to change over time. Page Rank has been defined by [Sergey Brin & Lawrence Page]: We assume page ‘A’ has pages T1, T2, T3,.............,Tn which point to it (i.e., are citations). The parameter ‘d’ is a damping factor, which can be set between 0 & 1. We usually set d 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: PR(A) = (1-d) + d (PR(T1)/C(T1) + ... + PR(Tn/C(Tn)) - (1). Page Rank can be thought of as a model of user behavior. We assume there is a “random surfer” who is given a Web page at random & keeps clicking on links, never hitting “back” but eventually gets bored & starts on another random page. The probability that the random surfer visits a page is its Page Rank & the damping factor d is the probability at each page the “random surfer” will get bored & request another random page. So it is very much important to add the damping factor d to a single page, or a group of pages to check Page Rank. This allows for personalization & can make it nearly impossible to deliberately mislead the system in order to get a higher ranking. The parameters C(A) is defined as the number of links going out of page ‘A’. So on the basis of these an another intuitive justification is that a page can have a high Page Rank if there are many pages that point to it, or if there are some pages that points to it & have a less Page Rank. As damping factor approaches
to ‘0’ the PR(A) from equation (1) becomes ‘0’. So Page Rank will be zero & Page ranking priority of that Web page will be at lowest level of hierarchy of WWW. Generally, the Page Ranks form a probability distribution over Web pages, so the sum of all Web pages’ Page Ranks will be one. A compromise between the two extremes of uniform [damping factor] & single page [damping factor] is to let [the damping factor apply from] all the root level pages of all Web servers. Aside from the Page Rank algorithm, Google also indexes anchor text (i.e. the text of a link), text proximity, & visual presentation details such as font size of words. These factors have several implications: Internal links on a site contribute to Page Rank, as well as links from other sites; a link from a home page is likely to contribute more to Page Rank than a link from other pages; the more links on a page, the less contribution each link will make to Page Rank; & the link text is significant.

There are two broad generalizations that can be made about search engine positioning: (1) word frequency in a page is important, particularly in the page title & near the top of the page. (2) The number of links to a page is important, particularly from sites that are in some way recognized as being authoritative & relevant. Of course, the question of what it is that makes a site authoritative or relevant, from the point of view of a search engine, is an interesting one. Different search engines address this issue in different ways. Other criteria that may be significant, include: text used in description & keywords meta tags; text used in links that point to a page; the proximity of words in the text for searches on phrases of two words or more; the size (and color) of font used for text; the use of H1-H6 tags; & whether a page is the root page of a domain. The generalized scripting codes for description, keywords, title, robots meta tags have been used for tourism based Website are as follows:

```html
<head><meta name="title" content="Department of Tourism"><meta name="description" content="An national tourism congress will be held in the month of Feb., 2006 at the Department of Tourism, Kurukshetra University, Kurukshetra, Haryana"><meta name="keywords" content="PPI, PPC, PPP & SEOs"><meta name="robots" content="ALL"><meta name="revisit-after" content="1 days"><meta http-equiv="refresh" content="8;URL=http://kuk.ernet.in"></head>
```

**B. Design criteria**

**B1. Usage of meta-tag**

Meta tags are used to define meta data. A meta tag carries information about a document & is placed in the HEAD of the HTML page. Meta tags are easy to implement on your Web pages. They look just like regular HTML tags, but they follow a few rules: (1) All begin with the word META, the element name, (2) All are located within the HEAD element, (3) No closing tag is used; the META element is an empty element. When you use a meta tag, you define attribute/value sets. For each attribute, you assign a value. If you look back at the element declaration, you'll see there are several available attributes, only one of which is required. So, the format might be diagrammed as: `<META attributeName="value" attributeName="value">`

**B2. Significance of meta tags in SEO**

Although there are no concrete rules defining what information a search engine spider pulls from your page, many search engines do make use of two primary meta tags: description & keywords. Both of these meta tags let you further define your site's content & function so that those searching for your site's content have a better chance of finding you & can get an accurate sense of what your site offers. Although some search engines index all the words on a Web page, often the first few words get the most exposure. For example, in the absence of a description meta tag, AltaVista uses the first few words on the page when it returns a set of search results. Using the description tag, you'll enter a sentence or two describing the page. Many search engines use this information when a list of search results is returned. The description tag for our gardening site might look like this: `<META name="description" content="Tourism Perspective in India: A case study">`. WebPromote suggests that you keep your description at less than 25 words. To get noticed, if you are tempted to put in a whole paragraph or more in description, don't waste the kilobytes. Search engines use a set amount of characters in the descriptions. If characters are too long, it will be cut off abruptly. WebPromote also suggests that you use lowercase keywords, as this is how most users enter search terms. After carefully considering the site's function & audience, the keywords tag for our tourism site might resemble the following: `<META name="keywords" content="tourism, hotels, supply chain management, agri products, gardening, garden, mountains">`. Not all search engines use all of these meta tags. E.g., Excite does not use meta keywords in indexing sites. It does, however, use the meta description for generating the summary of the site. They should be used to reiterate the words used in the title & description, including synonyms, plurals, regional variations & even common misspellings. In general: use all lower case; don't use commas & don't repeat words.

**B3. Use of Indexing & search phrases**

The over-enthusiastic use of search terms in your title, description & keywords tags is self-defeating & is likely to dilute the rank of your site; it is usually better for a page to appear near the top of one search query than for it to appear lower down on several. For this reason it is generally not recommended to target more than a single search phrase or, at most, two or three closely related phrases in the same page. Creating a robots text file lets you specify whether robots or spiders will index your site. But, you can also talk to the robots via the robots meta tag. This tag uses the name value robots & has two properties for content: index & follow. The following sample robots tag tells a) to not index this page, & b) to not follow links from this page to other pages: `<META name="robots" content="noindex, follow">`. The available values for each pair are: index or noindex; follow or nofollow.
B4. Layout of text & links on Web page

The first content that appears on every page should be a heading that reiterates the text of the title. This should be followed by a repeat of the description, & then the text that describes the main theme of the page. When linking to a page on your site, use the title of the page as the link text. For search engines that measure the relevance of links, this is likely to help reinforce the relevance of the page to the search terms used in the title. Try to avoid representing any of your link text as images. If you insist on it, then use Alt tags, so that engines can identify the text.

B5. Use of Grammatical aspects, frames & presentation elements on Web page

Some search engines distinguish between singular & plural versions of a word, while some don't, so, particular care needs to be taken where both versions are likely to be used in search terms. If both singular & plural versions of words are to be targeted, then there are two options: either creates two different pages that target the words separately; or, include both versions in the same page, which means including them both in the title & description of the page & using them both in your page copy. Be very careful when using frames. Some indexing agents will ignore your frameset & go straight to any text that has been placed outside the frameset. It is important to ensure that relevant text & links are placed here. These will need to be maintained & updated in tandem with the visible content on a site. Remember, even if you do manage to get your pages indexed, visitors to your site are likely to be referred to an orphaned frame lacking the navigational structure carefully constructed for it. Text contained in a Flash presentation will not be indexed by search engines. Flash should not be used for pages that you want to appear in search engines. Also, avoid excessive use of JavaScript. It makes a page harder to maintain & harder for search engines to decipher. Put any JavaScript functions in a separate file & link it from the page header.

C. Submission of Web page to directories & search engines

Before submitting your site to any directories, you should be satisfied that the title & description of your home page reflect your most important search terms & that they convey the right impression of your site. This should be the text that is submitted. By a wide & apparently increasing margin, Yahoo & Google are likely to be the most important in terms of generating traffic to your site. You should submit your home page to the major search engines. This should include, as a minimum: AltaVista, Excite, Fast, Google, HotBot, Lycos, & Northern Light.

D. Web Site architecture

While the keyword research is being done, your designer should be working on site templates & site architecture. Using Cascading Style Sheets (CSS), Frames, table tricks & external JavaScript will ensure that the search engine spiders are able to crawl your site without getting bogged down within the html code. You want them to be able to go to the page & quickly establish what your site is about. This will determine whether or not you get ranked. The content developer should work closely with your SEO analyst to make sure that the keywords are being used properly throughout the site. This means using heading (h1, h2, h3) tags, & links within the content using keywords, anchor text, etc. Finally, to make an effective web site, follow the basic ideas on web design techniques given in the web site of each engine's specific design guidelines. The some of the top most engines guidelines are as follows:

- Google says: (1) Make a site with a clear hierarchy & text links, (2) Offer a site map to your users with links that point to the important parts of your site, (3) Create a useful, information-rich site, & write pages that clearly & accurately describe your content (4) Think about the words users would type to find your pages, & make sure that your site actually includes those words within it (5) Try to use text instead of images to display important names, content, or links (6) Make sure that your Title & Alt tags are descriptive & accurate (7) Check for broken links & correct HTML.(8) If you decide to use dynamic pages, be aware that not every search engine spider crawls dynamic pages as well as static pages, (9) Keep the links on a given page to a reasonable number (fewer than 100).
- Yahoo! suggests: (1) Build a quality site, (2) Get big-sites with over 100 pages will attract more attention from search engines looking for sites with quality content, (3) Say something-sites with frequently updated & new content rank higher in search engines typically, (4) Use keywords not only as a tag, (5) Exchange links, (6) Get involved & get your name out there.

E. Clickthrough data in search engines

Clickthrough data in search engines can be thought of as triplets \((q,r,c)\) consisting of query \(q\), the ranking \(r\), & the set \(c\) of links user clicked on. E.g. the user asked query “India Tourism Perspective”, received the ranking shown in figure 5, & then clicked on the links ranked 1, 4, & 8 (marked in bold).

Since every query corresponds to one triplet, the amount of data that is potentially available is virtually unlimited. However, each query is assigned a unique ID, which is stored in the query log file of the server with the query words & the presented ranking. The links on the results page presented to the user do not directly suggest document, but point to a proxy server on which this document is stored. From the example of figure 5, it is not possible to infer that links 1, 4, & 8 are relevant on an absolute scale. Furthermore, it is much more probable to conclude that link 4 is more relevant than link 3 with probability higher than random [N. Fuhr 1989]. Assuming that the user scanned the ranking from top to bottom, he/she must have observed link 3 before clicking on 4, making a decision to not click on it. Given that the abstracts presented with the links are sufficiently
informative, this gives some indication of the user’s preferences. Similarly, it is not possible to infer that link 8 is more relevant than link 2, 3, 5, 6, & 7.

1. **Tourism Policy**
   www.equitabletourism.org/tourpolicy.htm - 34k - Cached - Similar pages

2. **Cooperatives and Tourism : An Asian Perspective**
   www.coop.org/ica/cartagenaverm.pdf - Similar pages

3. **Tourism and Sustainable Economic Development : Indian and Global**
   https://www.vedamsbooks.com/no41491.htm - 7k - Cached - Similar pages

4. **Development of Tourism and Travel Industry : An Indian Perspective**
   https://www.vedamsbooks.com/no43170.htm - 3k - Cached - Similar pages

5. **Indian Tourism in Perspective**

6. **Tourism of India - Special Feature - India Tourism Launches**
   www.tourismofindia.com/misc/marketing.htm - 16k - Cached - Similar pages

7. **Visa Southeast Asia | Travel and Tourism Reports**
   www.visa-asia.com/ap/sea/valueofvisa/industrywatch/vov_travel_and_tourisme.shtml - 35k - Similar pages

   www.researchandmarkets.com/reportinfo.asp?report_id=338113 - 42k - Cached - Similar pages

9. **India Tourism | East India - Excitement and Tranquility**
   www.india-tourism.com/EN/east_india_excitement_and_tranquility.html - 27k - Cached - Similar pages

10. **The Hindu : Tourism in social perspective**
   www.hinduonnet.com/2001/10/16/stories/1316017g.htm - 8k - Cached - Similar pages

---

**Figure 5: ranking obtained for the query “India Tourism Perspective”**

This means clickthrough data does not convey absolute relevance judgements, but partial relative judgements for the links the user browsed through. A search engine ranks the returned links according to their relevance to user’s query q, which ranked link 4 ahead of 3, & link 8 ahead of 2, 3, 5, 6, & 7. Denoting the ranking preferred by the user with $r^*$, we get partial relative information as follows:

$$\text{link}_4 < r^* \text{link}_5 \quad \ldots \ldots \ldots \quad \text{(2)}$$

$$\text{link}_8 < r^* \text{link}_2 \text{link}_6 < r^* \text{link}_5 \text{link}_9 < r^* \text{link}_3 \text{link}_8 < r^* \text{link}_6 \text{link}_9 < r^* \text{link}_7 \quad \ldots \ldots \ldots \quad \text{(3)}$$

Therefore, for a ranking $(\text{link}_1, \text{link}_2, \ldots, \text{link}_n)$ and a set $c$ containing the ranks of the clicked on links. We extract a preference equation: $\text{Link}_i < r^* \text{link}_j$ for all pairs $1 \leq j < i$.

V. **Future predictions & challenges**

For making the desired search of tourism information we are looking for non-textual information more heavily i.e. each and every user wants to look images, sounds, video etc. At present it is very difficult to search these information efficiently by search engines. Because the non-textual information requires Tera or Peta bytes of databases. In tourism based Websites the existence of non-textual information is more & therefore search engines should also handle the queries based on non-textual information in a precise & controlled manner. Near future will see the designing of non-textual based information retrieval system of tourism. Web search engines processed a user query on the basis of textual information. As far as tourism perspective is concerned, one should maintain the Websites with much more graphical coverage (such as images, program, & databases). So, the existence of images, programs & databases can’t be in cooperated by text-based search engines. Most search engines associate the text of a link with the pages that the link is on. In addition, we associate it with the page the link points to. There are many information retrieval challenges in addition to those briefly mentioned already in this paper, such as polysem, synonymy, query language, & speed. The Web presents its own unique challenges-- it can be viewed as one huge database with unique properties: large amounts of volatile data (rapid updates, broken links, file disappearances); heterogeneous data (multiple formats, languages, alphabets); an exponentially growing number of pages; redundant data; lack of structure; & a lack of an editorial review publication process, which leads to numerous information errors, a falsehoods, & invalid statements. Many search engines are beginning to incorporate structured text retrieval, meaning that users can search for text in italics, capitals, or boldface print. Google currently incorporates some structured retrieval. Users can choose to search for keywords located in the title, body, or hyperlinks of documents. Google's advanced features include searching for HTML documents, images, language filters, PostScript documents, or Excel documents, among other choices. Hotbot allows users to search for video, MP3-formatted data, or images, in addition to ordinary text. Its advanced features also include date information. E.g. users can choose to retrieve only recent documents, those updated in the last two weeks, two months, or two years. Another recent trend in search engine research is to use syntactic clues & query structure to improve retrieval results. These methods incorporate the location of the keywords in the query, the
use of prepositions, verbs, & adjectives, hyperlink structure on document pages, & citation links to enhance retrieval. The futuristic efforts will include the concepts of non-textual information system, pattern matching, fuzzy logic, & edit distance to eliminate & enhance the query results. In upcoming future, one of the most exciting applications of information retrieval will be an image retrieval system. True image retrieval uses mathematical models to capture how similar images are to an original image without using text hints, such as HTML image tags. Soon users will be able to retrieve images in addition to text documents. One application of image retrieval will be applicable for locating suspected criminals through a criminal photo IDs database electronically to find those criminals whose photo profiles most completely match the sketch rendered by the crime scene artist. But there are many parameters, which will be involved to search such type of non-textual information on the WWW. This can be further extended to sound retrieval systems, fingerprint information retrieval systems, & eye retina information retrieval systems. Searching through audio files is as challenging as image retrieval, but may be possible in the near future.

VI. References