

**Sixth International Conference on Webometrics, Informetrics and Scientometrics & Eleventh
COLLNET Meeting, October 19 – 22, 2010, University of Mysore,**

**ONLINE VISIBILITY OF WEBSITE THROUGH SEO TECHNIQUE:
A CASE STUDY**

Swastik Sinha ¹, Amritap Sinha ² and Swapan Deoghuria ^{3*}

**Indian Association for the Cultivation of Science
Kolkata – 700032, India**

Email: ¹ ccss@iacs.res.in, ² ccas@iacs.res.in, ³ ccsd@iacs.res.in

*** Author for correspondence**

Abstract: Search Engine Optimization (SEO) [2, 9, 11, 12] is widely used nowadays for improving the volume or quality of traffic to a website through search engines. The results generated by search engines can be natural (organic or algorithmic) and/or paid search. In this paper we have discussed different techniques used for achieving better optimization for the website of Indian Association for the Cultivation of Science (IACS), a premier research institute in India, devoted to the pursuit of fundamental research in the frontier areas of Physics, Chemistry, Biology, Energy, Polymer and Materials. Different techniques related to SEO like keyword discovery, crawling, on-page and off-page optimization and different Google tools [7, 8] are discussed. The results discussed in this paper and investigation will help webmasters to gain a deep insight about SEO and also to guide them in making better decisions regarding the online visibility of their websites.

Key Words: Search Engine Optimization, Online visibility, Webometrics, Web Crawler, Keyword Discovery.

1. Introduction

Online visibility of websites is very important for business, marketing and advertising. These are achieved by making and advertising good websites. But it is really difficult to advertise or sell ones' products if others do not find that particular website amongst the millions of websites available on Internet. Search engines are the most popular and dependent vehicle to find the required information and 85% web users use search engines [3, 6] by entering keywords in any of the search engines. Google is the most widely used search engine, where users can find information related to their interest.

Many people think that Search Engine Optimization (SEO) is quite expensive and time consuming. They prefer to go with paid advertisements on the Internet provided by different online companies or hire different companies to perform SEO campaign on their behalf. The main motivation behind this work is not only to optimize the website of IACS but also find a package that is suitable to most of the websites. Search for particular information on the Internet can be a very tedious task because of context. A search is dependent on many things and not easy to achieve. In order to get a desired goal or aim it is important to keep all these perspectives in mind [4].

IACS is one of the most prestigious research institutes in India where Sir C. V. Raman worked during 1907 to 1933 on various topics of Physics making discovery of the celebrated effect on scattering of light in 1928, which bears his name and that brought many accolades including the Nobel Prize in 1930. The American

Chemical Society designated the Raman Effect as an International Historic Chemical Landmark in 1998.

1.1 Aims and Objectives

The main aim of this work is to analyze, develop and optimize IACS website (<http://www.iacs.res.in>) using SEO techniques. Results can be helpful for researchers and practitioners who are working in the area of website design and development especially in SEO areas. Our goal was:

- To find new and existing techniques for SEO in order to apply on IACS website.
- To find tools that can be used in making websites more optimized in terms of SEO.
- Modifying the existing layouts of different WebPages of IACS including metadata insertion [15] to make them visible through keyword discovery.
- Research activities carried out at IACS and profiles of scientists working over here should be visible through keyword discovery.

Research Methodology

We used a case study approach that is a part of qualitative research. The focus was to search technical improvements in existing layouts of the pages using SEO standards. In this case study, we have investigated and evaluated different search engine optimization techniques in websites. This study also presents some useful techniques that can be used in any website without putting too much effort.

2. Background

Ideally everyone wants their sites to be more user-friendly and displayed somewhere on the first three pages of search results as most people won't look beyond the third page. The fact is, it is the sites that fall on the first page of results that get the most traffic, and traffic is translated into popularity or revenue, which is the ultimate goal of search engine optimization. A site must be recognizable by a search engine crawler to achieve a high position in search results. It must satisfy a set of criteria that not only gets the site catalogued, but can also get it catalogued above most of the other sites that fall into that category or topic.

Indian Association for the Cultivation of Science popularly known as IACS is the oldest research institute in India (established in 1876) and it is one of the most prestigious institutes doing research in frontier areas of basic sciences. IACS has its own registered domain and its website is accessible till April 1998. The website of IACS is developed and designed indigenously and the structure and contents of it changed continuously according to the requirement. The URL of IACS website was submitted to popular search engines immediately after its' first appearance. Although the keywords "Indian Association for the Cultivation of Science" or "IACS" worked fine and retrieved the official website of IACS, the homepages of scientists or their related pages could not be found by searching their names except a few renowned scientists holding positions like Director. As Webmaster of IACS website, we tried to find out the drawbacks of IACS website and wanted to make sure that all pages of IACS website including the homepages of scientists should be visible online through popular search engines by keyword discovery.

3. Search Engine Optimization (SEO)

SEO is a science of customizing elements of a website to achieve the best possible search engine ranking or appearance. In other words it aims to increase the number of visitors to a website by improving rankings [13]. There are many aspects of search engine optimization and its behaviour varies from website to website. Also, SEO depends on how the web pages of a website are designed. SEO essentially involves making it easy for search engines to find a website and boosting its position in their rankings. Basically, our objective is divided in to three major phases namely, planning, implementing and maintaining. These three phases are iterative and interrelated to each other. Following are the most important ingredients of SEO that we followed for IACS website.

3.1. Defining Goal

In this phase we carefully studied the target audience or/and visitors of IACS site. This helped in finding appropriate keywords and phrases to achieve our goal explained in section 3.3. A goal, in this context, defines a successful outcome from someone visiting our website and is expressed using a verb and a noun [14].

3.2. SWOT Analysis of similar Website

Main purpose of SWOT analysis in SEO is determination of websites current position and setting its perspective position in future together with website development strategy. We examined the websites of similar organizations for their look and features and their rankings in Search Engines.

3.3. Keyword Discovery

When a user visits a search engine, they type words into the search box to find what they are looking for. The search terms are also called keywords and the combinations of keywords are called key phrase. Keywords are not only most essential in making a website optimized [14] but also the most important and essential ingredients of SEO campaign. We invested significant time and energy on keyword research. It is observed that 20% of all SEO effort is focused on this activity alone [14] and poor keyword selections spoil its objective [14]. It is observed that 33% of all searches on search engines are for two-word combinations, 26% for three words, and 21% for four or more words. Just 20% of people search on single words [14]. Keyword analysis is done firmly based on objective facts about what people actually search on [14].

We made a scientific study of the keywords and key phrases our visitors and competitors actually use [14]. These keywords are used in Meta tagging described in section 3.6.3. Fig. 1 shows that 46.45% traffic sources are from search engines and 54.78% visits comes through the keyword “IACS”.

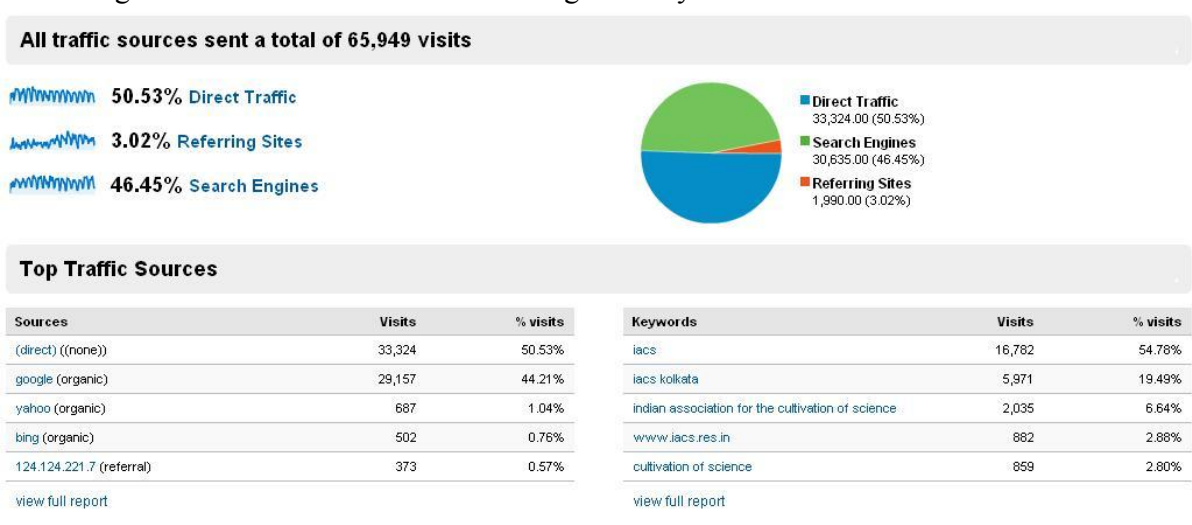


Fig.1

3.4. Crawling

Crawling is a mean by which a web crawler extracts or collects information about pages from a web site. This is all about helping Google to find and index our website. Google uses a spider program also know as robots or crawler to search web, collect documents, scan text and hand them over to indexing programme. Google’s spider is called GoogleBot and it maintains a log that one can see by searching. We used Google Webmaster for analyzing information about IACS web pages. We submitted our site to Google Webmaster, it crawls and extracts information from it and displays statistics accordingly. Robots.txt was also used because search engines other than Google are still using the old fashioned way of indexing [5]. Sitemap is very important in crawling and making pages more accessible. Detailed description about sitemaps is given in section 3.6.8. We have our own domain name that is used extensively in Meta tags. Site structure and navigation was done as much as possible as recommended by Google Webmaster [14]. Fig. 2 shows the crawl status of pages of IACS website during April to July, 2010 and it is evident that crawl started during April, 2010 after submitting Sitemap to Google and changed Robots.txt to allow crawling. After that, visibility of IACS website enhanced drastically and crawling of

pages is as high as 615 per day.

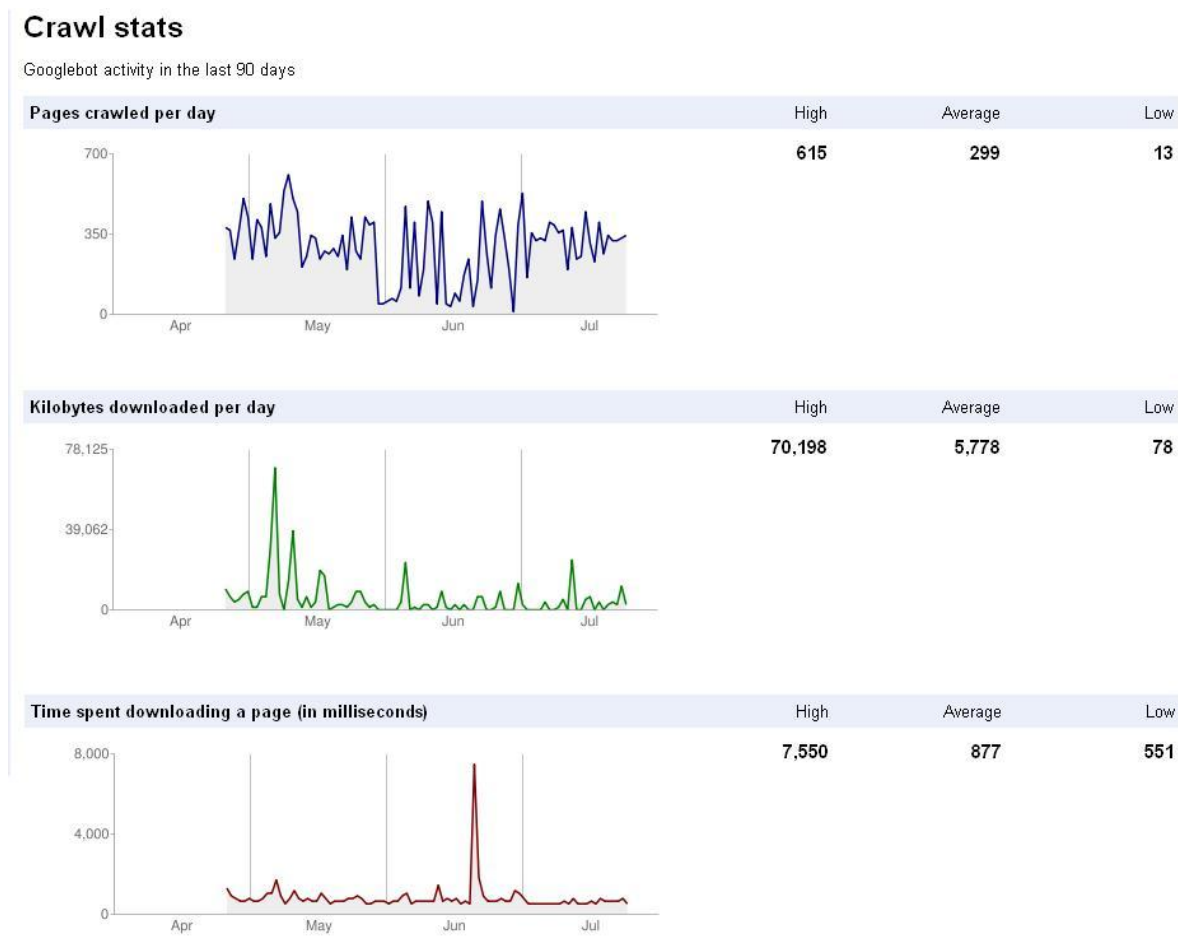


Fig. 2

3.5. Storing Information and Indexing Pages by Google

Google uses indexing mainly for optimizing, increasing the speed and performance of a search query by breaking down documents into keywords contained in it. This idea of indexing is taken from old card indexing system in libraries. While the use of an index speeds up information retrieval, the tradeoffs are the massive amount of additional storage required to hold the index and the considerable time and effort needed to keep that index up to date. Once GoogleBot has crawled, it gives a unique ID to each page it has found and passes these to an indexing program. This lists every document that contains a certain word [14]. **Fig. 3** shows that the keyword “IACS” retrieves not only the homepage of IACS but the related pages of the site also through Google.

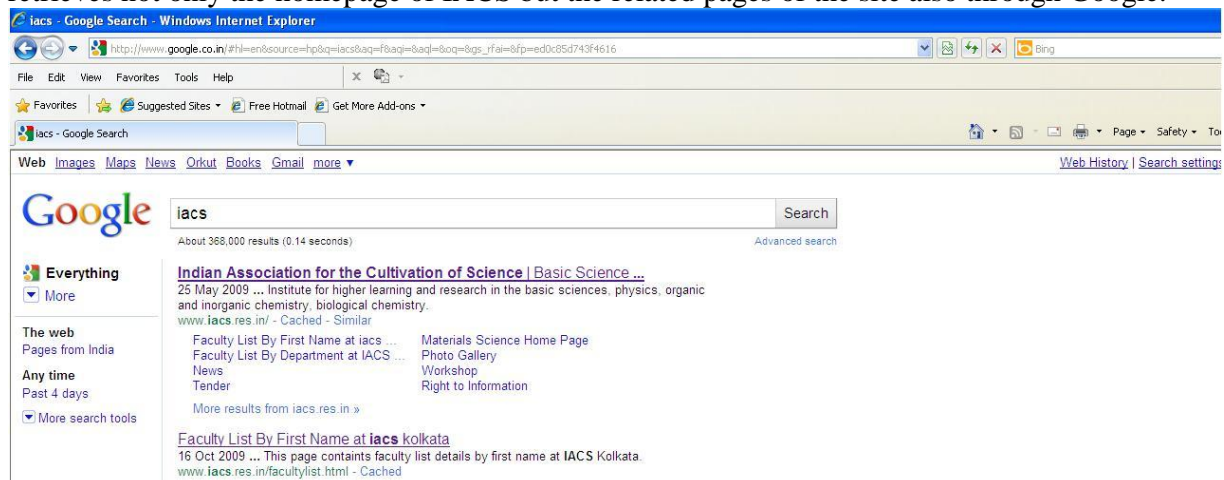


Fig. 3

3.6. On-Page Optimization

On-page elements like page titles, page metadata, headings, body text, internal links, and image alt tags etc. were studied extensively as given below:

3.6.1. Page Title

The page title tag tells both users and search engines about the topic of particular page. The page title is a string of text, defined by contents of the <title> element in the <head> section of the HTML document. The title is visible both in the title bar of a browser window, as well as the headline of a search engine result. It is arguably one of the most important factors in search engine optimization because it is both an important factor in search engine rankings, as well as a critical call to action that can enhance the click-through rate (CTR). This is important in SEO aspect because of **Algorithmic weighting, Results formatting, Browser usability and Directory submission** [14]. We used keywords in our page titles and tried to give unique titles for each and every page that increased the possibility of ranked high in Google. As per W3C guidelines we kept length of the title tags within 64 characters (including spaces) or less as far as possible. Search engine spiders use these title tags as the main source for determining the web-page topic. Qualitative improvement is achieved in finding each and every page in IACS website through search engine due to unique title of each page.

3.6.2. Meta Description

In Meta description we placed a brief description about our website build on the keywords selected earlier. Meta description is very important to improve sites CTR. We followed Google Webmaster tools guidelines, kept those as short and descriptive. 'Google Analytics' keyword statistics helped in the early stages of creating Meta description.

3.6.3. Meta Keyword

We used the meta-keyword tag in each and every page of IACS website that is placed between the <head> tags in HTML page and was intended solely for use by search engines. Meta tags are worth pursuing as Yahoo and Ask.com results are still influenced by them but it is almost useless for improving position on Google but in case of graphic intensive pages that have much of flash, images etc it is useful. Yahoo and Ask.com search engine always pay attention to keyword tags while Google's algorithm is complex and has lot of things to consider.

3.6.4. Meta Robots

Meta Robots tag is used to tell the search engine whether you want the web page indexed or not. It is recommended to use it only if one knows that his/her website would not be indexed. We did not use this tag in IACS website.

3.6.5. Meta Distribution

Meta distribution tag tells the search engine who the page is meant for and can be set to; global, for everyone, local, for regional sites, and UI, for Internal Use. We have used this for limited pages meant for internal use. The syntax is as follows:

```
<meta name="distribution" CONTENT="global">
```

3.6.6. Heading Tag

The heading tag is the next most important SEO on page element after <title> tag but it is not used extensively in IACS website.

3.6.7. Image alt Attribute

Images may seem like a straightforward component of a site, but we optimized them by using "alt" attribute that

provide information about the picture if a browser doesn't support images. Now almost all images in IACS website are searchable through Google Image Search including the pictures of scientists of IACS.

3.6.8. Sitemap

In general, there are two types of sitemaps. The first type of sitemap is a HTML page listing the pages of the site - often by section - and is meant to help users find the information they need.

XML Sitemaps - usually called Sitemaps, with a capital S - are a way to give Google information about any site. This is the type of sitemap used by Google for optimization while yahoo and other search engines use text based sitemaps.

We have created and submitted a Sitemap to Google so that it knows about all the pages on our site, including URLs that may not be discoverable by Google's normal crawling process.

Sitemap is important in making pages searchable. It can be in .txt or .xml form. For Google, XML-based sitemap is necessary while other search engines like Yahoo! use txt-based sitemaps. Sitemaps.org defines the standard protocol. There are four compulsory elements that the sitemap must have [14] that were followed in creating IACS sitemap. Google has its own sitemap generator but we created our own sitemap for convenience and flexibility. Google Sitemap Generator [1, 8] is a tool installed on a web server to generate the Sitemaps automatically. Unlike many other third party Sitemap generation tools, Google Sitemap Generator takes a different approach: it monitors web server traffic, and detect updates of a website automatically [8].

3.7 Off-Page Optimization

We did not use much off-page optimization like blogging etc.

3.8. Google SEO Tools

There are many Google SEO Tools but we used mainly Google Webmaster tools and Google Trends. We did not use Google AdWords and Google AdPlanner.

3.8.1. Google Webmaster Tool

It is a free tool for webmasters that helps webmasters better control how Google interacts with their websites and get useful information from Google about their site. We extensively used this service and solved the following [14]:

- Identified the problems of crawling by GoogleBot.
- Uploaded the XML Sitemap file.
- Took help to generate robots.txt file.
- Identified and solved issues with duplicate and problematic title and description Meta tags.
- Identified the keywords of top searches used to reach our site.
- Removed unwanted site links that Google used in results.
- Received notification of quality guideline violations and file for a site reconsideration.

Like Google, Yahoo has Yahoo! Site Explorer and Microsoft has Live Search Webmaster Tools. **Fig. 4** shows the Sitelinks created for IACS by Google Webmaster Tool.

Sitelinks

Sitelinks are links to a site's interior pages. Not all sites have sitelinks. Google generates these links automatically, but you can remove sitelinks you don't want.

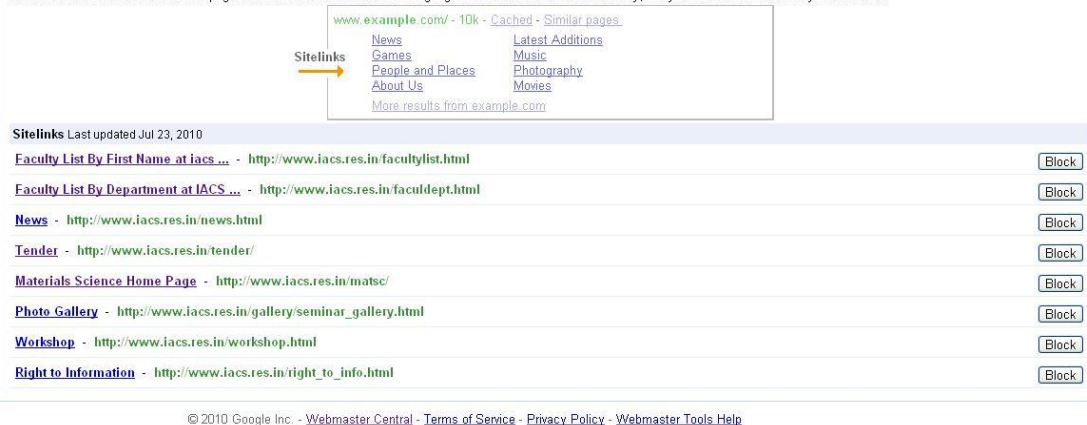


Fig. 4

3.8.2. Google Analytics Tool

After improving the crawling and indexing by using Google webmaster tools we looked for checking the incoming traffic to our site by Google analytics. Google Analytics helped us to understand

- How users reach and behave on our site.
- Helps in discovering the most popular content on our site.
- Helps in measuring the impact of optimizations we made to our site.

The **Table 1** shows the top five traffic sources and most used keywords used by visitors of IACS website

Sources	Visits	% visits	Keywords	Visits	% visits
Direct	33,324	50.53	iacs	16,782	54.78
Google	29,157	44.21	iacs kolkata	5,971	19.49
Yahoo	687	1.04	Indian association for the cultivation of science	2,035	6.64
Bing	502	0.76	www.iacs.res.in	882	2.88
124.124.221.7 (referral)	373	0.57	Cultivation of science	859	2.80

Table 1

The **Table 2** shows the top ten referring sites from where visitors reached IACS website

Source	Visits
124.124.221.7	373
google.co.in	223
apexicindia.com	207
dst.gov.in	149
twas.ictp.in	100
indiaedu.com	81
en.wikipedia.org	80
ispsquash.com	77
indiaonestop.com	70
goirectory.nic.in	67

Table 2

The **Table 3** explains the different search engines used by visitors to reach IACS site

Search Engine	Visits
google	29,157
yahoo	687
bing	502
search	235
ask	32
Bbaidu	19
aol	3

Table 3

3.8.3. Google AdWords

We did not use Google AdWords as IACS is not a profit making organization

3.8.4. Google Trends

We did not use much Google Trends in this study but we would use it to learn the statistics regarding the volume of keyword searches over various time periods in future.

3.8.5. Google Recommendations

There are different techniques and tools for achieving search engine optimization. Google has provided some tools for this purpose that now become a standard for every search engine optimization campaign. As Google is the most popular search engine, we used different Google SEO Tools like Google Webmaster Tool, Google Analytics Tool, Google Trends and Google Recommendations etc [3] for the same. Around January 2010 we created an account for Google Webmaster Tool for this purpose. First of all we submitted the sitemap of the website to Google for crawling as instructed. We selected homepages of ten scientists as test cases and studied the contents, key words, title and Meta tags and other features in detail. We found that in most cases these data were not unique. Contents of the pages were changed and proper title, Meta tags and unique keywords were inserted in the pages, trimmed javascript and inserted text based links in the pages as much as possible. The directory structure was also modified. After examining the crawl status, we found surprisingly that all links mentioned in the sitemap were not being crawled due to restrictions imposed in robots.txt file. We allowed the above mentioned ten pages as well as the sitemap file in robots.txt file. After a few days we found that those pages are searchable through keywords as inserted in those pages.

4. Results

After initial success with ten personal pages of scientists of IACS, we implemented the same with all pages of IACS website. The directory structure of WebPages was thoroughly modified to keep them crawl friendly. The sitemap was placed in the parent directory. Unique title was given to each page. Meta tags were checked and modified in favor of keyword discovery. We checked the common key words used by visitors to find IACS website through Google Analytics and implemented those keywords in related pages. We found remarkable improvements regarding online visibility of IACS till April 12, 2010. The pages crawled per day by Googlebot from IACS Website were around zero up to March 2010 whereas it reached a maximum of 919 in the last week of April, 2010 (**Fig 5**).

Crawl stats

Googlebot activity in the last 90 days

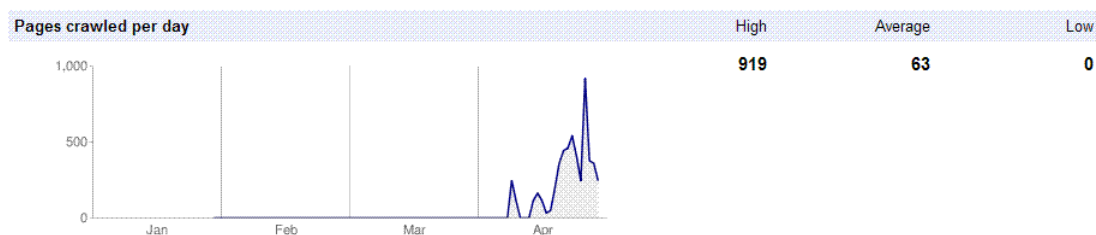


Fig. 5

The Fig. 6 shows the site performance of IACS website and it is evident that IACS site is faster than 68% of sites and it takes only 2 seconds to load for a visitor.

or tools



Fig. 6

5. Conclusion and future scope

Without going for paid search, only simple SEO techniques can help to make a website visible through keyword discovery in search engines. As on August 14, 2010 almost 1070 pages of IACS website are visible through Google search engine. All scientists are searchable through their names. Even their pictures and research drawings are searchable through Google image search. We checked other popular search engines like Yahoo, Bing, Ask.com etc and found favorable results. The entire study was carried out with Google SEO tools as we could not find a common tool in one place that is compatible with every search engine. In future this study can be extended to include other SEO tools and make a comparative study. Also Web 2.0 features are not considered in this study and we think RSS [10] and other features of Web 2.0 would be useful in this type of study.

References:

1. About sitemaps. <http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=156184>. Retrieved on 14-08-2010
2. Advantages of SEO. <http://tusharvickkie.blogspot.com/2008/10/advantages-of-seo.html>. Retrieved on 14-08-2010
3. Anatomy of Search results. <http://googlewebmastercentral.blogspot.com/2007/11/anatomy-of-search-result.html>. Retrieved on 14-08-2010
4. Andrew Goodman. "Search Engine Showdown: Black hats vs. White hats at SES".

<http://searchenginewatch.com/showPage.html?page=3483941>. Retrieved on 14-08-2010.

5. Brian Pinkerton. "Finding What People Want: Experiences with the WebCrawler" (PDF). The Second International WWW Conference Chicago, USA, October 17–20, 1994.

http://www.webir.org/resources/phd/pinkerton_2000.pdf.

6. Georgia_Tech "How Users Find out About WWW Pages," 1998, http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/graphs/use/q52.htm. Retrieved on 14-08-2010

7. Google WebMaster guideline for optimization of your website.

<http://www.google.com/support/webmaster/bin/answers.py?answer=35769>.

<http://www.google.com/webmaster/guidelines.html>. Retrieved on 14-08-2010

8. Google Sitemap Generator <http://googlesitemapgenerator.googlecode.com/svn/trunk/doc/gsg-intro.html>.

Retrieved 14-06-2010

9. Ledford, Jerri. L. Search Engine Optimization Bible. Wiley Publishing Inc. 2008. ISBN: 978-0-470-17500-2

10. Richardson, Will. 2005. The ABCs of RSS. *Technology & Learning*, 25 (May 2005), Issue 10, 20-24

11. Search Engine Optimization. http://en.wikipedia.org/wiki/Search_engine_optimization. Retrieved on 14-08-2010.

12. Search Engine optimization (SEO).

<http://google.com/support/webmasters/bin/answer.py?hl=en&answer=35291>. Retrieved on 14-08-2010

13. The basics of search engine optimization. <http://blogscoped.com/archive/2007-01-07-n13.html>. Retrieved on 14-08-2010

14. Viney, David. Get to the Top on Google. Nicholas Brealey Publishing. 2008. ISBN: 978-1-85788-502-6

15. Zhang, J. and Dimitroff, A. 2005. The impact of metadata implementation on webpage visibility in search engine results (part II). *Inf. Process. Manage.* 41, 3 (May. 2005), 691-715.