Submitted on:27.06.2016

Report of publishing open National Bibliography of Iran

Saeedeh Akbari-Daryan
Advisor to the Deputy Research, Planning and Technology, National Library and Archives of Iran, Assistant Professor of the National Library and Archives of Iran, Tehran, Islamic Republic of Iran.
sakbaridaryan@gmail.com

Hassan Bagheri Mahabadi
The project manager of Integrated Library Systems , National Library and Archives of Iran. The Bachelor of Arts in Computer Science, Tehran, Islamic Republic of Iran.
h_bagheri@hotmail.com

Fariborz Khosravi
Deputy Research, Planning and Technology, National Library and Archives of Iran. Assistant Professor of the National Library and Archives of Iran, Tehran, Islamic Republic of Iran.
fa.khosravi@gmail.com

Abstract:

Providing and upgrading national bibliographies is an expensive process, while in recent years, the behaviour of library users has been undergoing change, as they rely increasingly on web search engines, rather than libraries, to meet their information needs. Considering the lack of appropriate infrastructure and necessary tools in different countries, the object of this study is to find the shortest and fastest way for publishing the open national bibliography in search engines. This Study discusses about how NLAI OPAC is submitted in Google Webmaster and how it is monitoring.

Keywords: Open National Bibliography of Iran (NBI), Search engines, Google, Search Engine Optimization (SEO), Google Search Console (Google Webmaster Tools), OPAC. National Library and archives of Iran (NLAI)

INTRODUCTION

"A current national bibliography is a mirror that reflects the culture of a country. By looking at the current national bibliography one is able to learn about the uniqueness of a country. The emphasis on agriculture and technology, the make-up of its society through its various language publications, particular customs and ceremonies important in the life of the nation, the importance of education, literature and science, prominent literary authors of the time and

political, social and religious trends within a country are all discernible. A current national bibliography should reflect the interests and unique characteristics of a country much as a mirror reflects the uniqueness of an individual" (Best Practice for National Bibliographic Agencies in a Digital Age, 2014).

National library and Archives of the Islamic Republic of Iran (NLAI) is an educational, scholarly, research, and service-providing institution which operates under the supervision of a Board of Trustees headed by the President of the Islamic Republic of Iran. The official establishment and inauguration of National Library goes back to 1937 and its constitution has been approved by National Consultative Assembly in 1991. NLAI consists of four deputies: 1. National Library 2. National Archives 3. Research, Planning, and Information Technology 4. Administrative that operates in two separate buildings: National Library building, and National Archives building. (Shakeri, Akbari-Daryan & Mohammadi, 2012).

NLAI first published National Bibliography of Iran (NBI) in 1963. The CD-ROM of NBI was presented in 1996 .The NLAI OPAC system was set on the web in 2006. At present, NBI can be used either on-site or through remote access (PC, mobile phone and other remote devices) by NLAI users.

NLAI policies on scope of NBI are as follows:

- All textual and non-textual materials which are published inside the country including books, periodicals, pamphlets, newsletters, posters, maps, audio and video tapes and discs, electronic and digital documents etc., through the National Deposit Act or purchase;
- Materials published outside the country by Iranians;
- Manuscripts existing inside the country regardless of language (Teymourikhani, Akbari-Daryan, 2013).

Providing and upgrading national bibliographies is an expensive process, while in recent years, the behaviour of library users has been undergoing change, as they rely increasingly on web search engines, rather than libraries, to meet their information needs. Search engines have become an integral part of our information environment. Increasingly they are replacing the role of libraries in facilitating information discovery and access. Googling has become synonymous with research. Recent statistics indicate that Google has become the search interface of choice for many faculty and students to address their information needs, far exceeding their use of library catalogues or other online citation databases (Rieger, 2009). "According to the OCLC reports in 2006, 89% of the undergraduate and graduate students surveyed start their search for information with web search engines, and only 2% start at the library website. Other statistical data support this trend" (Sadeh, 2007).

Search Engine Optimization (SEO), or search engine positioning, is the process of identifying factors in a webpage, which would affect search engine accessibility to it, and fine-tuning the many elements of a website so it can achieve the highest possible visibility when a search engine responds to a relevant query. Search engine optimization aims at achieving good search engine accessibility for Web pages, high visibility in search engine results, and improvement of the chances the Web pages are retrieved (Gasparotto, 2014).

Google now processes over 40,000 search queries every second on average which translates to over 3.5 billion searches per day and 1.2 trillion searches per year worldwide (Internet Live Stats, 2016).

According to the statistics recorded in the Google, 48.9 percent of Iranian population is Internet users (30,749,524 people). Statistics and reports of NLAI OPAC (2011 – 2015) show that the averages of daily and annual processing searches are 9,762 and 3,563,036 respectively. However, despite the lack of funds, the NLAI costs a lot of money to produce the national bibliography.

The fact is that the NLAI should adopt a strategy to increase the visibility of its website in search engines essentially, so the authors are encouraged to study and find solutions to increase the visibility of NLAI OPAC. In the beginning study phase, authors considered Linked Data option.

Linked Data provide a way for libraries to enhance their visibility through SEO. Exposing library metadata as Linked Data would mean it could be crawled by the search engine bots and included in the search results presented to users. Linked Open Data (LOD) projects are happening all around the world, expanding the way that we access cultural heritage. Many of the projects of LOD are still at a proposal stage. The important challenge confronting potential adopters is the complexity of Linked Data technologies such as RDF/XML, RDFS, OWL and SPARQL. " Linked Data applications will not matter if search engines cannot first find library websites and repositories, crawl them, and understand the metadata provided" (Arlitsch, 2015).

Therefore, the authors of this paper aim to find ways and means to publish open national bibliographies other than Linked Data due to the aforementioned challenges. Considering the lack of appropriate infrastructure and necessary tools in different countries, the object of this study is to find the shortest and fastest way for publishing the open national bibliography in search engines. This study describes our experiences in developing and implementation of optimization project for NLAI OPAC.

Methodology

The method of this research is combination of the library research and trial and error method. At the conclusion of the first Phase i.e. study phase, researchers chose Google Search Console.

Google Search Console (previously Google Webmaster Tools) is a no-charge web service by Google for webmasters. It allows webmasters to check indexing status and optimize visibility of their websites. Google Search Console is a free service offered by Google that helps libraries to monitor and maintain their website's presence in Google Search results. Google Webmaster Tools provides the libraries with the top search queries their websites appears.

After selecting Google Webmaster tools in the 1st phase (study phase), the authors have defined the rest of the project in below phases:

2nd phase: investigating the process and trend of submitting websites in Google Webmaster Tools.

3rd Phase: Submitting NLAI OPAC and monitoring the progress process.

4th phase: Analysing registered errors in Google Webmaster.

5th Phase: Increasing the quality of pages according to registered errors.

6th phase: analysing and comparing searching status of NLAI OPAC before and after optimization.

As you know SEO is a time-consuming process and noting the limited time for preparing the project process to IFLA 2016, this study has gone through the first to Third phases.

After selecting Google Webmaster in the 1st phase, in the 2nd phase submitting website has been considered. This was done by rigorous study on *Search Engine Optimization: Starter Guide* (Search Engine Optimization: Starter Guide, 2010). 3rd phase is started in 18 April 2016 with submitting 306,424 URLs in 11 files. According to the progress of the indexing project in 22 April 529,286 URLs was submitted again in 19 files. Table 1 depicts the status of submitted files and indexed URLs.

Table 1 shows that from 835,710 web pages submitted 5,315 web pages are indexed. This figure is till 16 May 2016 that shows within one month, 0.6 of submitted pages are indexed. This process has been with 7 warnings that all contained the same message:

> *Some URLs listed in this Sitemap have a high response time. This may indicate a problem with your server or with the content of the page.*

Noting the received messages, the contents of the files that got warning, were analysed again. Apparently comparing to other files, some files had high response time that Google would consider on as a problem. The contents of these files were checked and nothing wrong with them was found. In order to further investigation, *Index Status Report* was surveyed.

TABLE 1- SITEMAPS OF SEARCH CONSOLE (16 MAY 2016)

Sitemap	Processed	Submitted	Submitted Web	Indexed Web
http://opac.nlai.ir/sitemap_00	2016/05/12	2016/04/18	27853	183
http://opac.nlai.ir/sitemap_01	2016/05/09	2016/04/18	27857	189
http://opac.nlai.ir/sitemap_02	2016/05/15	2016/04/18	27857	199
http://opac.nlai.ir/sitemap_03	2016/05/11	2016/04/18	27857	196
http://opac.nlai.ir/sitemap_04	2016/05/10	2016/04/18	27858	203
http://opac.nlai.ir/sitemap_05	2016/05/12	2016/04/18	27857	203
http://opac.nlai.ir/sitemap_06	2016/05/13	2016/04/18	27857	167
http://opac.nlai.ir/sitemap_07	2016/05/15	2016/04/18	27857	204

Sitemap	Processed	Submitted	Submitted Web	Indexed Web
http://opac.nlai.ir/sitemap_08	2016/05/14	2016/04/18	27857	195
http://opac.nlai.ir/sitemap_09	2016/05/15	2016/04/18	27857	210
http://opac.nlai.ir/sitemap_10	2016/05/14	2016/04/18	27857	174
http://opac.nlai.ir/sitemap_11	2016/05/16	2016/04/22	27858	153
http://opac.nlai.ir/sitemap_12	2016/05/15	2016/04/22	27857	132
http://opac.nlai.ir/sitemap_13	2016/05/15	2016/04/22	27857	139
http://opac.nlai.ir/sitemap_14	2016/05/14	2016/04/22	27857	154
http://opac.nlai.ir/sitemap_15	2016/05/13	2016/04/22	27857	193
http://opac.nlai.ir/sitemap_16	2016/05/14	2016/04/22	27857	181
http://opac.nlai.ir/sitemap_17	2016/05/14	2016/04/22	27857	172
http://opac.nlai.ir/sitemap_18	2016/05/15	2016/04/22	27858	121
http://opac.nlai.ir/sitemap_19	2016/05/15	2016/04/22	27857	108
http://opac.nlai.ir/sitemap_20	2016/05/12	2016/04/22	27857	186
http://opac.nlai.ir/sitemap_21	2016/05/13	2016/04/22	27857	183
http://opac.nlai.ir/sitemap_22	2016/05/15	2016/04/22	27857	109
http://opac.nlai.ir/sitemap_23	2016/05/16	2016/04/22	27857	165
http://opac.nlai.ir/sitemap_24	2016/05/13	2016/04/22	27857	205
http://opac.nlai.ir/sitemap_25	2016/05/15	2016/04/22	27858	197
http://opac.nlai.ir/sitemap_26	2016/05/15	2016/04/22	27857	209
http://opac.nlai.ir/sitemap_27	2016/05/12	2016/04/22	27857	193
http://opac.nlai.ir/sitemap_28	2016/05/12	2016/04/22	27857	211
http://opac.nlai.ir/sitemap_29	2016/05/15	2016/04/22	27857	181
Total			**835710**	**5315**

Index Status Report

The Index Status Report provides data about the URLs that Google tried to index in the current property for the past year.

This Report includes parts below:

- **Total indexed URLs in your site**

This part shows the total URLs available to appear in search results, along with other URLs Google might discover by other means. This number changes over time as you add and remove pages. The number of indexed URLs is almost always significantly smaller than the number of crawled URLs, because Total indexed excludes URLs identified as duplicates, non-canonical, or those that contain a Meta no index tag.

- **URLs blocked by robots.txt**

The total number of URLs disallowed from crawling by your robots.txt file. If your site is very big, you might want to hide other data so that the graph is scaled to a readable range.

- **URLs removed**

The number of URLs you have removed with the URL removal tool. Again, this value should be quite low in comparison to the other URLs in this report, so it's easier to view this selection by itself rather than in comparison with other URLs (Search Console Help, 2016).

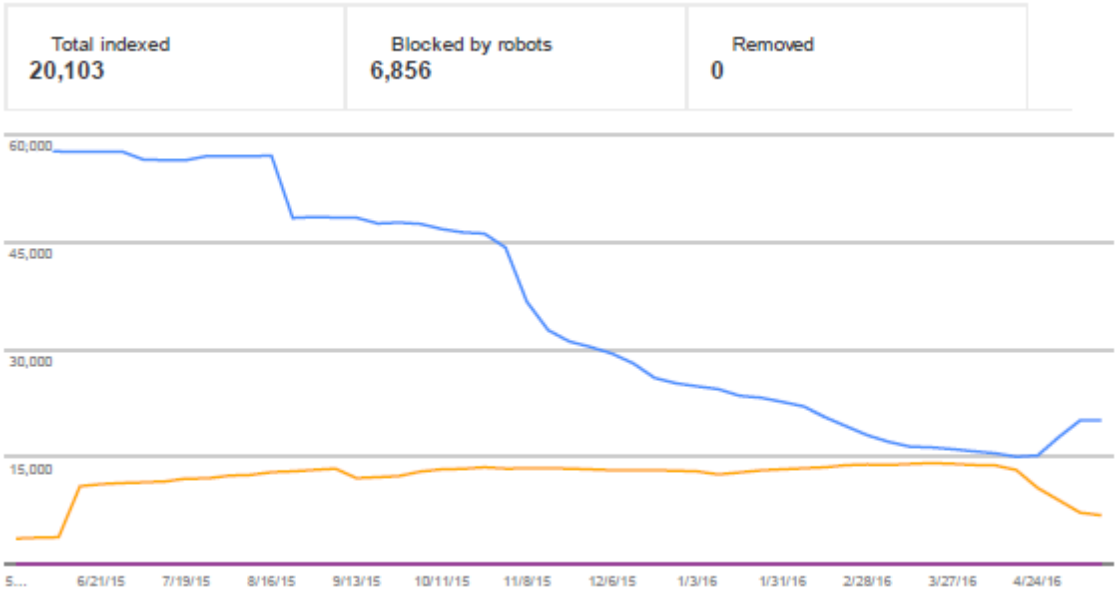

CHART 1- INDEX STATUS REPORT OF NLAI OPAC (16 MAY 2016)

Chart 1 shows that the index has an upward trend at first, but it has been decreased in last week seriously. The blocked URLs in the Chart 1 were related to the past security policies opened in this step. Chart 2 shows the results after opening all blocked URLs.

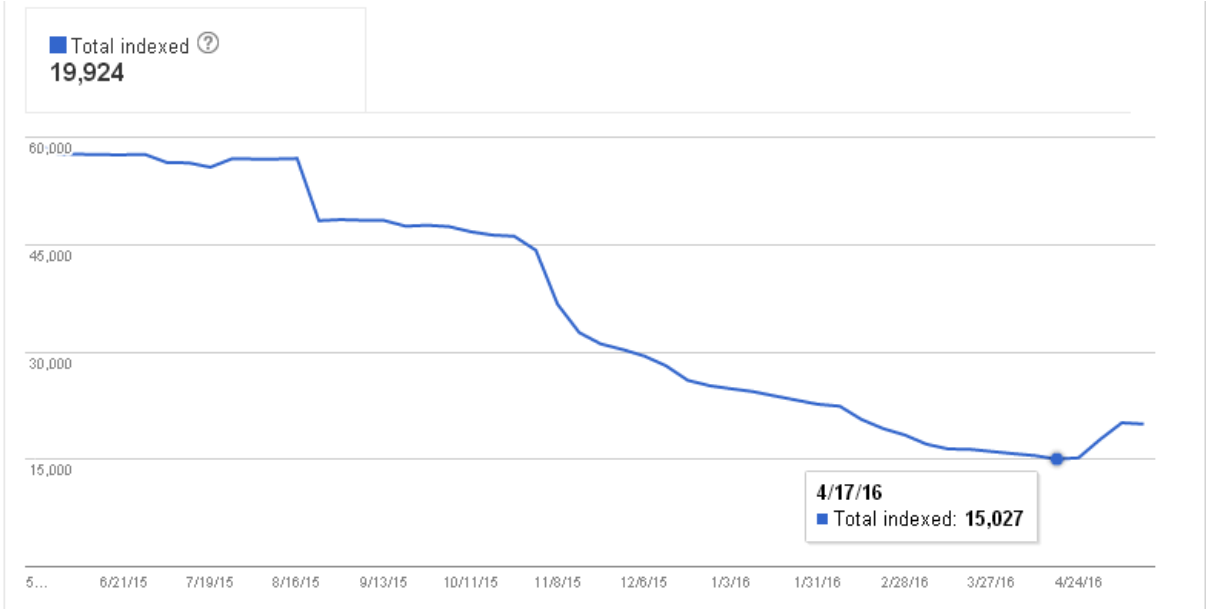

CHART 2- INDEX STATUS REPORT OF NLAI OPAC (21 MAY 2016)

In Chart 2 we can see the reduction of Total indexed comparing to Chart 1. Therefore *Crawling Activity and Errors* was monitored in two separate sections:1) *Crawl Errors Report* 2) *Crawl Stats Report*.

1. Crawl Errors Report

The Crawl Errors Report for websites provides details about the site URLs that Google could not successfully crawl or that returned an HTTP error code. The report has two main sections:

1. Site Errors: This section of the report shows the main issues for the past 90 days that prevented Googlebot from accessing your entire site.
2. URL Errors: This section lists specific errors Google encountered when trying crawl specific desktop or phone pages. Each main section in the URL errors reports corresponds to the different crawling mechanisms Google uses to access your pages, and the errors listed are specific to those kinds of pages (Search Console Help, 2016).

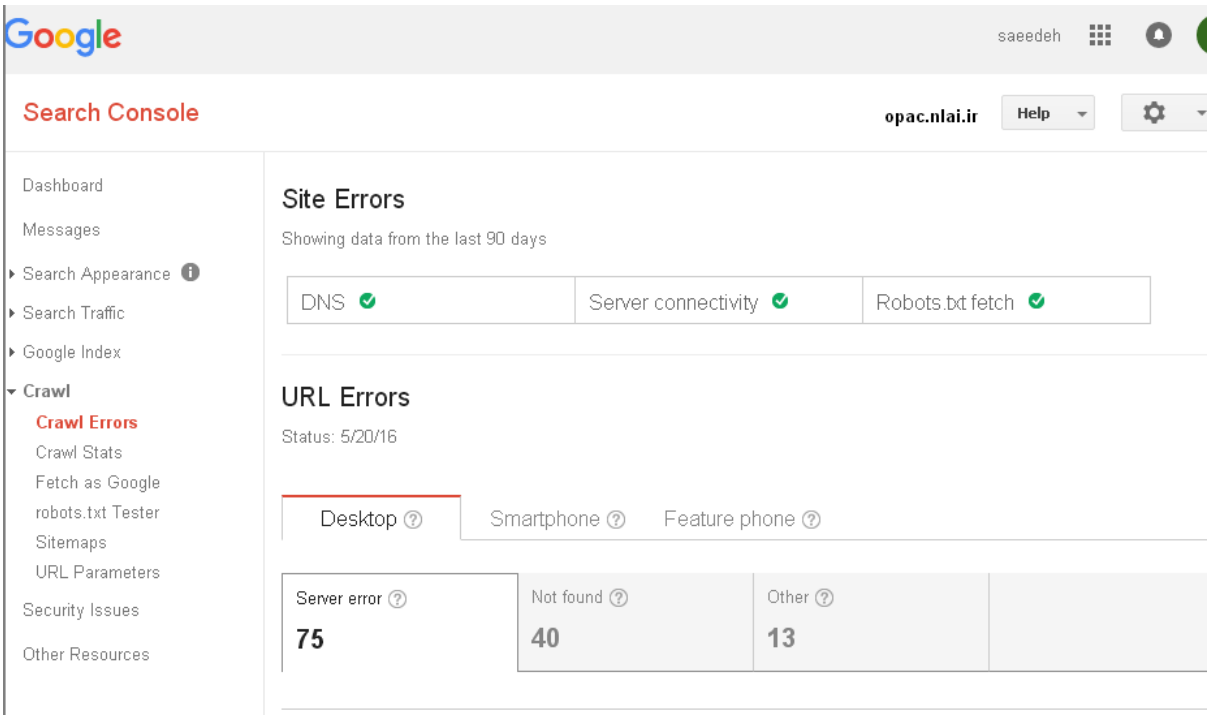

FIG. 1- CRAWL ERRORS REPORT OF NLAI OPAC (20 MAY 2016)

As you can see in Figure 1 the Crawl Errors Report displays the errors in two separate parts: Site errors and URL errors but All Crawl Errors of NLAI OPAC are related to the URL errors part.

Although the Search Console Help notifies not every error you see in this section - URLs Errors- requires attention on your part, but it's important that you monitor this section for errors that can have negative impacts on your users and on Google crawlers.

TABLE 2- URL ERRORS OF NLAI OPAC (20 MAY 2016)

URL errors	Number of errors	HTTP error code
Server error	75	500
Not found	40	404
other	8	400
total	**123**	

As you see in table 2 the highest rate of errors -HTTP error code 500- is the generic error message, given when an unexpected condition was encountered and no more specific message is suitable. It should be noted that because of time limitations to send this report to the IFLA, The third phase of the project began early. This phase coincide with NLAI topology change project of network infrastructure and configuration of network schedule. Therefore it was expected we would receive http error code 500 in this period.

2. Crawl Stats Report

The Crawl Stats Report (for websites only) provides information on Googlebot's activity on your site for the last 90 days. These stats take into account all content types that we download.

There's no "good" crawl number, but you should see a relatively even chart that increases over time as you increase the size of your site.

Chart 3 shows that number of pages crawled per day are 5,086 averagely.

CHART 3- THE NUMBER OF PAGES CRAWLED PER DAY IN NLAI OPAC (16 MAY 2016)

According to Chart 4 the average data downloaded per day in NLAI OPAC is 38,883 Kilobytes.

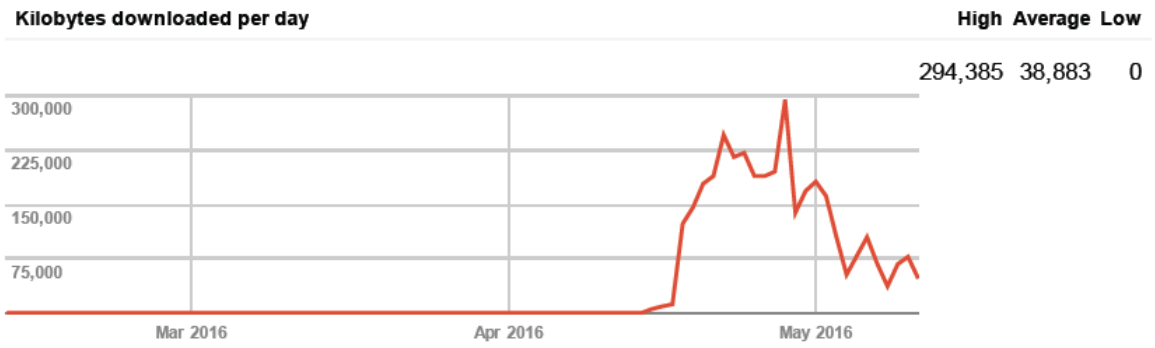

CHART 4- THE AMOUNT OF DATA DOWNLOADED PER DAY IN NLAI OPAC (16 MAY 2016)

Chart 5 shows that the time spent downloading a page in NLAI OPAC is 2 milliseconds averagely.

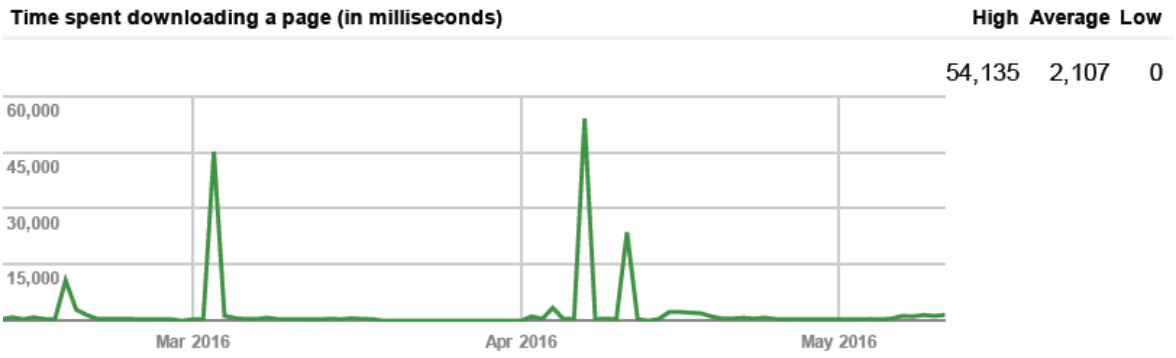

CHART 5- THE TIME SPEND DOWNLOADING A PAGE IN NLAI OPAC

The last three charts show that data crawled has increased and the time spend downloading a page compared with the time before Google Indexed Pages, has decreased.

HTML IMPROVEMENTS REPORT

One of the important points in monitoring recorded errors is detailed analysis and subsequently improving the quality of data in websites indexed. This type of analysis will be done in the 4th phase of the project.

The HTML Improvements page shows you potential issues Google found when crawling and indexing your site. We recommend that you review this report regularly to identify changes that potentially increase your rankings in Google search results pages while providing a better experience for your readers.

These issues don't prevent your site from being crawled or indexed, but paying attention to them can improve the user experience and even help drive traffic to your site.

In this section, Google webmaster Recorded 449 *Duplicate Title Tags* in16 may 2016 and 465 items in 21 may 2016 for NLAI OPAC. Investigating some samples of Duplicate Title Tags indicates some of them are not URLs submitted by OPAC's webmaster. It seems they include pages that Google encountered them in following internal links and some other include items that Google has found as Duplicate titles but they are different in other information. For example maybe there are two books with the same title but they are not the same in publishers, date of publishing, etc. This part needs investigating all items one by one that will be done in 4th phase.

Search Analytics Report

The Search Analytics Report is an added value that helps you to know how often your site appears in Google search results. Use the results to improve your site's search performance . The following metrics are available:

- **Clicks** - Count of clicks from a Google search results page that landed the user on your property.
- **Impressions** - How many links to your site a user saw on Google search results, even if the link was not scrolled into view? However, if a user views only page 1 and the link is on page 2, the impression is not counted. The count is aggregated by site or page. With infinitely scrolling pages, such as image search, the impression might require the item to be scrolled into view.

- **CTR** - Click-through rate: the click count divided by the impression count. If a row of data has no impressions, the CTR will be shown as a dash (-) because CTR would be division by zero.
- **Position** - The average position of the *topmost* result from your site. So, for example, if your site has three results at positions 2, 4, and 6, the position is reported as 2. If a second query returned results at positions 3, 5, and 9, your average position would be (2 + 3)/2 = 2.5. If a row of data has no impressions, the position will be shown as a dash (-), because the position doesn't exist.

TABLE 3- SEARCH ANALYTICS OF NLAI OPAC

Devices	Clicks	Impressions	CTR	Position
Desktop	11484	88226	13.02%	5.5
Mobile	2954	42345	6.98%	6.2
Tablet	398	4855	8.20%	6
Total	**14836**	**135426**	**10.96%**	**5.9 (average)**

Table 3 shows that from 135,456 NLAI OPAC links that users have seen at Google results 14,836 items have led click on link. This number is so little and should consider specifically in 4th phase. In other words, in the best case i.e. use from desktop by users, observation of 13 percent NLAI OPAC in Google results has led to go to that page. However, average results position of OPAC in Google was 5.9. So in order to increase ranking should consider special measures in 4th phase. Of course, issue of Unclicking on OPAC's links in case of foreign speakers is justifiable, since Persian language is not useful for them. One of the effective measures can be increasing English metadata in indexed pages. For example, in regards of analytic metadata of National Bibliography of Iran, adding English equivalent of subject Headings and descriptors is possible in show page. In fact, List of Persian subject headings is bilingual and cultural thesaurus used in National Bibliography of Iran is three lingual (Persian, English, Arabic). Russian language and Tajik script are added to it.

More study in analytical Google shows that in case of US users CRT is 8.88 and in case of Russia and Iraq are 5.73 and 4.47 respectively. So this issue confirms necessity of adding metadata into other languages. In other words, in US, 9 present of observing links were clicked in results of Google searching.

In regards of aimed OPAC users, Persian speakers, and reasons of decreasing click on searched results should survey more.

Mobile Usability Report

The mobile usability report identifies pages on your site with usability problems for visitors on mobile devices.

TABLE 4- MOBILE USABILITY REPORT

Usability issue	Pages with issues
Viewport not set	654
Text too small to read	654
Clickable elements too close together	654
Content wider than screen	101

Table 4 shows that mobile's users have encountered more problems in use of NLAI OPAC. It is true that NLAI OPAC is not designed for mobile users at first, but in this case in next phase of project should use available solutions.

Despite of overview of recorded errors by Google, reason of slow indexing in last week is not specified. This issue will be considered more in 4th phase of project.

Conclusion

It seems that if libraries use search engine optimization tools, library websites will appear much more in research result. It is more important and vital about national libraries that have a large valuable data, while search engines don't include added value of authorities and references. But if OPACs can index their pages using search engines, they can benefit great search engines capacities in order to enriching OPACs. Authors believe that optimization project is of great value in library websites and libraries should take the project serious and increase their presence and visibility in Google using its tools.

References

- Arlitsch, Kenning (2015). Walk Before You Run Prerequisites to Linked Data Dean of the Library. Accessed may 21, 2016 https://www.oclc.org/content/dam/oclc/events/2015/CI_SFPL_PPTS/Arlitsch-OCLC-NCTPG_2015-02-10.pdf

- Best Practice for National Bibliographic Agencies in a Digital Age, (2014). Accessed may 21, 2016 http://www.ifla.org/node/5226

- Gasparotto, M. (2014). Search Engine Optimization for the Research Librarian: A Case Study Using the Bibliography of US Latina Lesbian History and Culture. *Practical Academic Librarianship: The International Journal of the SLA Academic Division*, 4(1), 15-34. Accessed may 21, 2016 https://journals.tdl.org/pal/index.php/pal/article/view/6971/6076

- Internet Live Stats.(2016) Accessed may 21, 2016 http://www.internetlivestats.com/

- Rieger, O. Y. (2009). Search engine use behavior of students and faculty: User perceptions and implications for future research. *First Monday*, 14(12). Accessed may 21, 2016 http://journals.uic.edu/ojs/index.php/fm/article/view/2716/2385

- Sadeh, T. (2007). User-centric solutions for scholarly research in the library. *Liber Quarterly*, 17 (3-4). Accessed may 21, 2016 https://www.liberquarterly.eu/articles/10.18352/lq.7897/

- Search Console Help (2016). Accessed may 21, 2016 https://support.google.com/webmasters

- Search Engine Optimization: Starter Guide (2010). Accessed may 21, 2016 http://static.googleusercontent.com/media/www.google.com/en//webmasters/docs/search-engine-optimization-starter-guide.pdf

- Shakeri, Shakeri, Akbari-Daryan, Saeedeh, & Mohammadi, Farnaz (2012). Comparison of Traditional and Digital Reference Questions and Responses at the National Library and Archives of IR of Iran (NLAI). *Reference & User Services Quarterly*, 52(2), 136-144. Retrieved Accessed may 21, 2016 https://www.questia.com/library/journal/1G1-318105236/comparison-of-traditional-and-digital-reference-questions

- Teymourikhani, Afsaneh & Akbari-Daryan, Saeedeh (2013). Future National Bibliography of the Islamic Republic of Iran (NBI): the users' perspective. Accessed may 21, 2016 http://library.ifla.org/124/