

Solving SEO Issues in DSpace-based Digital Repositories

A Case Study and Assessment of Worldwide Repositories

Matúš Formanek

ABSTRACT

This paper discusses the importance of search engine optimization (SEO) for digital repositories. We first describe the importance of SEO in the academic environment. Online systems, such as institutional digital repositories, are established and used to disseminate scientific information. Next, we present a case study of our own institution's DSpace repository, performing several SEO tests and identifying the potential SEO issues through a group of three independent audit tools. In this case study, we attempt to resolve most of the SEO problems that appeared within our research and propose solutions to them. After making the necessary adjustments, we were able to improve the quality of SEO variables by more than 59% compared to the non-optimized state (a fresh installation of DSpace). Finally, we apply the same software audit tools to a sample of global institutional repositories also based on DSpace. In the discussion, we compare the SEO sample results with the average score of the semi-optimized DSpace repository (from the case study) and make conclusions.

INTRODUCTION AND STATE OF ART

Search engine optimization (SEO) is a crucial part of the academic electronic environment. All their users attempt to process too much information and need to retrieve information fast and effectively. Making academic information findable is essential. Digital institutional repository systems, used to disseminate scientific information, must present their content in ways that make it easy for researchers elsewhere to find. In this paper, we describe work conducted in the Department of Mediamatics and Cultural Heritage at Faculty of Humanities, University of Zilina to improve the discoverability of materials contained within its DSpace institutional repository.

In the literature review, we examine definitions of website quality and discuss audit tools. Then, beginning our case study, we describe the tools applied at our institution. We next describe the selection process of a suitable set of testing tools, focused on the optimization of SEO variables of the selected institutional repository running with DSpace software, that will be applied later in the case study. The remainder of the article focuses on the identification and resolution of potential SEO issues using the three independent online tools we selected. We aim to resolve as many problems as possible and compare the level of achieved improvement with the default installation of DSpace 6.3 software which our digital repository is based on. The primary goal is not only to improve the SEO parameters of the discussed system but also to increase the searchability of scientific website content disseminated by DSpace-based digital repositories.

Next, we offer insights into worldwide DSpace-based repositories. We will show that DSpace is currently one of the most widely used software packages to support and run digital repositories. Unfortunately, there are many major SEO issues that will be discussed later. The secondary objective of this paper is to use the same set of tools to evaluate the current state of the sample of worldwide digital repositories also based on DSpace. We will provide the report based on our own findings. In the discussion, the SEO score of the optimized DSpace (from the case study) will be

Matúš Formanek (matus.formanek@fhv.uniza.sk) is Assistant Professor in the Department of Mediamatics and Cultural Heritage, Faculty of Humanities, University of Zilina, Slovakia. © 2021.

compared with the results of the current state of SEO parameters from the worldwide DSpace repositories.

Finally, our work also carries out many relatively innovative approaches related to digital repositories that have not been extensively debated anywhere in the literature yet.

LITERATURE REVIEW

To achieve our goal, we started with a review of existing academic papers. Drawing from those papers we describe the current state of academic institutions' presentation through the Internet and search engines. In this sense, we focus on website optimization.

The Internet, as a medium, is still rapidly expanding. A massive amount of data is communicated, shared, and available online, as noted by Christos Ziakos:

As a result, billions of websites were created, which made it hard for the average (or even advanced) user to extract useful information from the web efficiently for a specific search. The need for an easier, more efficient way to search for information led to the development of search engines. Gradually, search engines began to assess the relevance of every website on their indexes compared to the queries provided to them by the users. They took into consideration several website characteristics and metrics and calculated the value of each website using complex algorithms. The enormous number of websites being indexed from search engines, along with the increasing competition for the first search results, led to studying and implementing various techniques in order for websites to appear more valuable in search engines.¹

That description applies equally to academic websites as well as commercial ones. A review of relevant literature suggests that it is very important for academic institutions to carefully consider and apply website optimization. There were around 28,000 universities worldwide in 2010, according to one study that monitored research in the field of worldwide academic webometrics.² The actual number of universities seems to be very similar in 2020.

Baka and Leyni affirm in their working paper that the success or failure of an academic institution depends on its website: "The work of each university exists only when it encounters and interacts with society. Their popularity with the public is steadily growing." What is directly connected with the institution's presence in the World Wide Web.³

Many authors define the term search engine optimization (SEO) as a series of processes that are conducted systematically to improve the volume and quality of traffic from search engines to a specific site by utilizing the working mechanism or algorithm of the search engine. It is a technique of optimization a website's structure and content to achieve a higher position in search results. The aim is to make increase the website's ranking in a web search results.⁴

After an extensive information retrieval in the relevant literature, we can conclude that although SEO is currently a widely discussed topic, there is very little accessible scientific literature related to SEO applications in the field of digital repositories in general, and none at all in the particular subset of DSpace-based repositories.

Website Quality

Many authors generally affirm that there is a positive correlation between academic excellence and the complex web presence of an institution. It confirms that website quality is a factor that can give us a predictive or causal relationship with SEO performance.⁵ Numerous tools could be employed to measure the quality of websites, test them closely and produce an SEO performance ranking websites ability to properly promote their content through the search engines. For example, the Academic Ranking of World Universities (The Shanghai Ranking, <http://www.shanghairanking.com>) has been established for the top 1,000 universities in the world. The website quality is considered by the authors as the quality of institution's online presence, its ability to properly promote digital content in search engines and finally, in combination, its overall web presence. According to the Shanghai Ranking list, this is a factor for some "prospective students to decide on whether they will enroll in a specific institute or not."⁶ A number of recent studies have also attempted to examine the online presence of academic institutions from various points of view. One of the older studies mentioned that the quality of academic websites is very important for students in the process of enrollment.⁷ Another key aspect is the optimized website performance as well as SEO and website security.⁸

Audit Tools

If we want to perform any optimization, we need an appropriate software tool to check a current website's ranking. According to G2, the world's largest technology online marketplace, SEO software is designed to improve the ranking of websites in search engine results pages without paying the search engine provider for placement. These tools provide SEO insights to companies through a variety of different features, helping identify the best strategies to improve a website's search relevance.⁹ SEO audit software could be used by SEO specialists or system administrators, as well. Audit software performs one or more of the following functions in relation to SEO: content optimization, keyword research, rank tracking, link building, or backlink monitoring. The software then provides reports on the optimization-related metrics.¹⁰ Many authors stress the importance of a holistic approach to SEO factors (24 factors were tested), but it depends on the most effective ones: for example, the quantity and quality of the backlinks, the SSL certificate and so on, which will be described later in this paper.¹¹

The quality of academic websites is very important for researchers, too. They need to disseminate scientific information and communicate it in effective ways. According to some authors, the topic of academic SEO (ASEO) has been gaining attention in recent years.¹² ASEO applies SEO principles to the search for academic documents in academic search engines such as Google Scholar and Microsoft Academic. In another scientific paper, ASEO is considered as very similar to traditional SEO, where institutions want to make good use of a SEO to promote digital scientific content on the Internet. Beel, Gipp, and Wilde emphasize the importance for researchers to ensure that their publications will receive a high rank on academic search engines.¹³ By making good use of ASEO, researchers will have a higher chance of improving the visibility of their publications and have their work read and cited by more researchers.

In recent years, digital institutional repositories (as the academic systems) have been used as modern ways of promotion and dissemination of digital scientific objects through the Internet.

Digital objects need to reach a wider audience—digital repositories have a form of website interface, interact with students, teachers, or researchers on a daily basis and use the number of citations, articles, theses or other research objects. Institutional repositories are affected by search

engines too, so some improvements on repositories' SEO parameters are needed. These factors contribute to a system's rankings.

SEO on institutional repositories is not considered an absolutely new scientific topic. Kelly stressed eight years ago that Google is critical in driving traffic to repositories. He analyzed results from a survey describing the summarization of SEO findings for the 24 institutional repositories in the United Kingdom. The survey results showed that referring platforms were primarily responsible for driving traffic to those institutional repositories—thanks to many hypertext links in referring domains.¹⁴ Since then, SEO analyses of digital repositories have not been a widely discussed topic in the literature. It is a relatively unique topic to discuss SEO on a specific type of digital repository software—DSpace, as the most used and popular software for running digital libraries and repositories.¹⁵ Consequently, this paper focuses on that topic since the DSpace-based digital repository is a complex online computer system where some SEO parameters could be adjusted. SEO audit tools help to identify areas of potential adjustments of those website properties that could help produce higher rankings in search engines (and improve the whole system visibility).

AUDIT TOOLS SELECTION PROCESS

Website variables that affect SEO can be tested using specialized online software tools. This topic is discussed in detail on a semi-professional level on specialized websites that provide a number of recommendations regarding the use of specific tools as well as evaluations of the tools.¹⁶ These tools can keep track of changes in many SEO variables. We want to use this approach in our study. However, first we need to choose the appropriate set of these tools.

We have found that many SEO audit tools mentioned in professional online sources are narrowly specialized.¹⁷ For example, they may be focused only on keyword analysis, backlink analysis (for example, Ahrefs' Free Backlink Checker), and so on. In our study, we intend to describe a greater number of SEO parameters to monitor rather than emphasize only a few selected ones. We also need tools that are fully available online for free. Based on these criteria, we immediately excluded several tools from the selection, because they provide only austere, simple, or restricted information. Many tools were excluded because they were limited to a single test with the requirement of registration or provision of an email address. A number of testing tools were also available only in paid versions. We wanted a set of tools that focus on several aspects of SEO analyses and evaluate the quality of websites' SEO variables comprehensively. It is important to add that the selected tools results must be comparable, too.

After careful consideration of all possibilities, we finally decided to choose three independent SEO audit tools in order to make the approach more transparent. The selected tools met most of the criteria mentioned above. However, it is very important to note that many other software tools surely meet the criteria and could also be suitable for testing purposes. Based on the scientific literature review, we were not able to identify specific recommendations in this regard; therefore, we have been inspired by the advice offered in the websites and blogs previously mentioned that are focused primarily on SEO. Our tools selection is as follows (listed in alphabetical order):

1. **SEO Checker** (<https://suite.seotesteronline.com/seo-checker>) is part of a complex audit software suite called SEO Tester Online Suite. SEO Checker provides tests in the following categories: base, content, speed, and connections to social media. It tracks, among many other parameters, title coherence, text/code ratio, accessibility of microdata, OpenGraph

metadata, social plugins, in-page and off-page links, quality of links, mobile friendliness of the page and many other SEO and technical website attributes.

Regarding restrictions, only two sites can be tested within a 24-hour period. The limit increases to four sites per day after free registration with a valid email address. Moreover, there is a 14-day trial period during which all hidden functionalities work. In the free version that we used, a complete report can be viewed only, not downloaded or saved.

2. **SEO Site Checkup** (<https://seositecheckup.com/>) was selected based on many positive recommendations from the technically oriented expert website Traffic Radius.¹⁸ SEO Site Checkup is described as “a great SEO tool that offers more than 40 checks in 6 different categories (common SEO issues like missing metadata, keywords, issues related with absence of connections to social media, semantic web, etc.) to serve up a comprehensive report that you can use to improve results and the website’s organic traffic. It also gives recommendations to fix critical issues in just a few minutes. As a tool, it is very fast and provides in-depth information about the various SEO opportunities and accurate results.”¹⁹

SEO Site Checkup is appreciated and recognized as number one among other audit tools ranked by the Geekflare website.²⁰ Another reason we selected this tool for our testing scenario is the fact that the Google search engine will offer a link to this tool as the first after entry the search query “seo testing tool” (excluding paid links). SEO Site Checkup is also the fastest of the selected audit tools, which could be considered as another advantage. Its disadvantages include the ability to test only one website within 24 hours from one public IP address.

3. **WooRank** (<https://woorank.com>) is recommended by Traffic Radius: “WooRank offers an in-depth analysis that covers the performance of existing SEO strategies, social media and more. The comprehensive report analysis is classified into eight sections for improved readability quotient, and you may also download the report as branded PDF.”²¹ WooRank has obtained the third position among the recommended software tools. TrustRadius gives it a score of 9.2 out of 10 and users rate it of 4.67 out of 5 stars based on 51 reviews.²²

On the one hand, some results are hidden in the free version, but the final score will be shown. On the other hand, WooRank has no limit to the number of websites tested per day, but it is the slowest of the selected testing tools.

We selected these three SEO audit tools because they work independently, their results are comparable to each other, and they offer a quick way to get comprehensive SEO analysis results for a tested site. It should be noted that results of some performed tests are hidden, but there is general guidance on how to fix some issues. However, the solution always depends on the specific site and used technology.

Using three different tools adds objectivity because we do not rely on just one tool and a one-sided view of the SEO issue. The three selected testers all display results in the same way—test results are always shown as a summarized score in the range of 0 to 100 points (100 represents the best result). A very large set of SEO parameters and technical website properties is evaluated in all three cases. These tests are usually divided into several categories (for example, common SEO issues, performance, security issues, and social media integration). Although similar parameters

are assessed in all three audit tools, there are still some differences between them. Each of the testing tools is unique in a certain area because it also tests a parameter that the others do not deal with or evaluates a website by a different methodology. Still, the fact remains that the evaluated SEO parameters overlap between the tools.

We will not overload this paper with detailed information and technical details of individual partial tests, because they can be easily found on the website of the given test tools (SEO Site Checkup, SEO Checker Online, WooRank). We will just mention the common core of main tests: CSS Minification test, Favicon test, Google Search Results Preview test, Google Analytics test, H1 Heading Tags test, HTML Page Size test, Image Alt test, JavaScript Minification Test, JavaScript Error Test, Keywords Usage Test, Meta Description Test, Meta Title test, SEO friendly URL test, Sitemap test, Social Media test, Robots.txt test, URL Canonicalization test, and URL Redirects test.

Another specific group consists of tests related to a particular audit tool. Thanks to them we can get a more comprehensive view of the tested area of a website's SEO characteristics. For example, SEO Checker features the following specific tests: Title Coherence test, Unique Key Words test, H1 Coherence test, H2 Heading Tags test and Facebook Popularity test. WooRank as the-second tool extends the basic set of tests with the following: Title tag length test, In-page links test, Off-page links test, Language test, Twitter account test, Instagram account test, Traffic estimations and Traffic rank. Of course, there is also a set of tests that are parts of two audit tools, but the third one does not deal with them since it is specialized in another area.

As we have mentioned, the tools offer a list of suggestions for potential improvement of SEO characteristics. The user is informed about an issue, but no instructions or solutions are provided on how to resolve it. The main benefit of this paper lies with its objective to solve specific SEO issues. This work may improve the visibility and searchability of DSpace-based institutional repositories.

A set of the three audit tools described above will be used in the following section. We attempt to identify possible SEO issues of the selected institutional repository in the form of a case study. Then we aim to fix the identified SEO issues and increase its quality of SEO parameters as well as demonstrate the potential impact on website traffic caused by performed repairs. All traffic measurements will be based on Google analytics data.

THE INSTITUTIONAL REPOSITORY OF THE DEPARTMENT OF MEDIAMATICS AND CULTURAL HERITAGE (SEO CASE STUDY)

Background Information

An older version of our digital repository (based on DSpace v5.5) was launched by the Department of Cultural Heritage and Mediamatics in April 2017. Now, in 2021, the repository makes available online over 180 digital objects, most of them open access under Creative Commons licenses. The first attempts to create and establish a similar virtual space for digital objects started long ago. Several software solutions had been tested for this purpose—for example, Invenio and Eprints, along with DSpace. According to OpenDOAR's statistics, Eprints and DSpace have always been the most popular tools for running digital repositories.²³

A few years ago, DSpace was chosen as the primary software for running a digital repository. Since then, the usage of open-source software has been raising. For example, Ubuntu server LTS (long term support) is used as an operating system, Tomcat 8 is used as a web server, PostgreSQL

assumes the role of a database system, etc. All of those software components are part of a complex digital system and are orchestrated in a virtual environment that is built on an open-source virtualization solution called XCP-ng (in version 8.2). Some software components have been switched for others during the development period.

Based on our experience, the digital repository's regular visitors were mostly from the staff and students of the department. We initially did not feel a need to improve the visibility of this system to search engines, an oversight that turned out to be a mistake in the long run. We did not perform any search engine optimization on this repository until November 2019, when we coincidentally discovered several scientific articles dealing with SEO in the academic environment. After studying the theoretical background, we initiated the practical application process. We applied theory and our experience with DSpace software into an SEO troubleshooting process within our local repository. Most of the optimizing actions related to solving the major SEO issues were performed before November 10, 2019. We will describe the SEO adjustments we made and derive a list of recommendations for other institutions based on our own experience.

Initial Testing of a Clean DSpace 6 Installation

In order to formulate any recommendations related to SEO and the administration of DSpace digital repositories, it is important to determine and test a starting point. For this purpose, we chose a clean instance of DSpace v6.3 with an XML user interface (XMLUI)—the latest commonly available stable version. This is the same version that we use in this case study and in our production environment. (A newer version, DSpace 7 Beta 4, was released by Atmire on October 13, 2020).²⁴ No other customization edits were made except a base configuration and necessary URL settings. This installation of DSpace v6.3 has been tested by the same set of tools mentioned previously.

The tests we performed are summarized in table 1, where they are divided into four main SEO sections in the first column: common SEO issues, social, speed and security. A test name is shown in the second column. The third column is marked as "Default installation," where we display the test results on our clean DSpace 6.3 installation. If the tested instance met the criteria of the given test, the green pictogram occurs. When the particular test fails, the red cross is used.

The improved state is shown in the fourth column marked as semi-optimized. It is a consequence caused by many important technical changes and SEO issues solving process. This issue will be discussed and described later in this paper; however, a short note about the considered issue is displayed in each row. These notes were retrieved by reports on results.

We have used the prefix semi- in the last column because we were not able to resolve all detected SEO issues—only most of them. All related reasons will be described briefly in the discussion section. When the improving change between states has been made, we have changed a status pictogram (from the red cross to the green correct tick) and set the row color to yellow. The changes leading to improvement (e.g., the yellow rows) will be discussed in detail later, too.

Recall that we have no need to overload the main text of this paper with detailed technical information about partial tests, because it can be easily found on the websites of the given test tools.

Table 1 shows the compared results between the non-optimized and semi-optimized states of the DSpace repository. Based on table 1, the default instance of DSpace with basic HTTP and other

default settings received only 58 points out of 100 in SEO Site Checkup, 50.1 points in SEO Checker and 32 points in WooRank. The average final score is 46.7 points out of 100. Although this gained score could be considered as low, the DSpace default instance still meets certain basic criteria of SEO. In addition, many repository administrators usually do not rely only on a default installation, but they make at least some changes in configuration immediately after the initial installation. Inter alia, the first thing to do should be an implementation of HTTPS protocol, adding a connection with Google analytics services and so on.

The improved state is shown in the last column of table 1. Whenever we solved an issue, the overall score raised. The semi-optimized repository has obtained a higher score compared to the previous column (default installation). The last column represents the final (however semi-optimized) state of technical and SEO attributes which we were able to reach at this moment. As shown, many SEO issues have been solved. We highlighted them in yellow. On the one hand, some issues remain unsolved. On the other hand, the overall SEO improvement is more than noticeable although the final average gained score has not reached the maximum value (100 points).

Table 1. Comparison of results between the non-optimized and semi-optimized states of DSpace repository.

Test name	State	
	Default installation (before optimization)	Semi-optimized (after a few optimization steps)
Meta Title test, Title tag length	 The title tag is set, but the meta title of the webpage (DSpace Home) has a length of 11 characters. It is too low.	 The title tag has been set to "Digitálny repozitár Katedry mediamatiky a kultúrneho dedičstva" (note: in Slovak language).
Title coherence test	 The keywords in the title tag are included in the body of the page	 The title of the page seems optimized.
Meta Description test	 No Meta-description tag is set.	 Meta-description tag has been set. (121 characters)
Google Search Results Preview test	 "DSpace Home" is too general.	 The title of the page has been changed.
Keywords Usage test	 The keywords are not included in Title and Meta-description tags.	 A set of appropriate keywords has been added.
Unique key words test	 The textual content is not optimized on the page.	 There is an excellent concentration of keywords in the page. This page includes 382 words of which 58 are unique.
H1 Heading Tags test	 8 H1 tags, 6 H2 tags	 The H1 tags of the page seem not to be optimized. There are too many H1 tags.
H1 Coherence test	 The keywords present in the tag h1 are included in the body of the page.	 Some of the keywords of the tag h1 are not included in the body of the page.
H2 Heading Tags Test	 The keywords present in the tag <h1> are included in the body of page.	

Test name	State	
	Default installation (before optimization)	Semi-optimized (after a few optimization steps)
Language test	 Detected: Slovak Declared: Missing	 A missed language tag has been implemented.
Robots.txt test	 No "robots.txt" file has been found.	 "Robots.txt" file has been enabled.
Sitemap test	 No sitemap has been found.	 Sitemap has been enabled.
SEO friendly URL test	 Webpage contains URLs that are not SEO friendly!	 Webpage contains URLs that are not SEO friendly.
Image Alt test	 The webpage does not use "img" tags. It is optimized.	
Inline CSS test	 The webpage uses inline CSS styles.	 The webpage uses inline CSS styles.
Deprecated HTML Tags test	 The webpage does not use HTML deprecated tags.	
Google Analytics (GA) test	 GA is not in use.	 GA has been implemented.
Favicon test	 Default DSpace favicon is used.	 The favicon has been customized.
JS Error test	 No severe JavaScript errors were detected.	 No severe JavaScript errors were detected.
Social Media test	 No connection with social media has been detected.	 The website is successfully connected with social media (using Facebook).
Facebook account test	 	 Information about Facebook page has been added by schema.org metadata.
Facebook popularity	 (low)	 The webpage is promoted enough on Facebook.

Test name	State	
	Default installation (before optimization)	Semi-optimized (after a few optimization steps)
Twitter account test	 No connection with Twitter has been detected.	 Information about Twitter account has been added by schema.org metadata.
Twittercard test	 No twittercard is implemented.	 Metainformation about twittercard has been added by OpenGraph metadata.
Instagram account test	 No connection with Instagram has been detected.	 Information about Instagram account has been added by schema.org metadata.
Microdata (OpenGraph, Schema.org) test	 There is no microdata or OpenGraph/schema.org metadata on the website.	 Some OpenGraph and schema.org metadata has been added.
HTML Page Size test	 The size of the page is excellent. (23.65 KB)	 The size of the page is excellent. (28.84 KB)
Text/code ratio test	 10.71% (excellent)	 15.45% (excellent)
HTML Compression/GZIP	 (no compression is enabled) The size of HTML could be reduced up to 79%.	 The webpage is successfully compressed using gzip compression on your code. Your HTML is compressed with 78% size savings.
Site Loading Speed test	 Loading time is around 1.86s	 Loading time is around 2.39s
Page Objects test	 The webpage has fewer than 20 http requests.	 The webpage has fewer than 20 http requests.
Page Cache test (server-side caching)	 The pages are not cached.	 The pages are not cached.
Flash test	 Website does not include flash objects.	

Test name	State	
	Default installation (before optimization)	Semi-optimized (after a few optimization steps)
CDN Usage test	 Your webpage is not serving all resources (images, javascript and css) from CDNs.	 Your webpage is not serving all resources (images, javascript and css) from CDNs.
Image, Javascript, CSS Caching tests	 Data are not cached.	 Data are not cached.
Javascript Minification test	 Javascrpts are not minified.	 JavaScript files' minification has been enabled in Tomcat configuration.
CSS Minification test	 Some of your webpage's CSS resources are not minified.	 Some of your webpage's CSS resources are not minified.
Nested Tables test	 The webpage does not use nested tables.	
Frameset test	 The webpage does not use frames.	
Doctype test	 The website has a valid doctype declaration.	
URL redirects test	 1 URL redirect has been detected. It is acceptable.	
URL Canonicalization test	 The webpage URLs are not canonized.	 https://repozitar.kmkd.uniza.sk/xmlui and https://www.repozitar.kmkd.uniza.sk/xmlui should resolve to the same URL, but currently do not.
Canonical Tag test	 No canonical tag has been detected.	 The webpage is using a canonical link tag.
HTTPS test	 Website is not SSL secured.	 HTTPS has been implemented.

Test name	State	
	Default installation (before optimization)	Semi-optimized (after a few optimization steps)
Safe Browsing test	✓ No malware or phishing activity found.	✓
Server signature test	✓ Server self-signature for HTTPS is off.	✓
Directory Browsing test	✓ Server has disabled directory browsing.	✓
Plaintext Emails test	✓ The webpage does not include email addresses in plain text.	✓
Mobile friendliness (includes tap targets, no plugin content, font size legibility, mobile viewport)	✓ The webpage is optimized for mobile visitors.	✓
SEO Site Checkup final score	58/100	81/100
SEO Checker online final score	50.1/100	78.0/100
WooRank final score	32/100	65/100
Average final score	46.7/100	74.66/100

Resolving Major SEO Issues

This section will look at how we resolved the major SEO issues that the tools detected. This is the key technical part because most of mentioned issues highlighted in table 1 were solved and described. The following technical and SEO adjustments have been implemented and tested in order to improve the average final score by 59.87% (by 27.96 points, from 46.7 to 74.66 points)—comparing the fresh installation of DSpace against the semi-optimized one.

All the following solution procedures are based on our own experience, experiments, and research carried out in the area of digital repositories and their optimization as virtual spaces. During the solving process, we follow the order of issues stated in table 1 and describe them in more details in the DSpace v6.3 environment and an XML user interface (XMLUI). The following procedures may differ slightly if you are using a different version of DSpace or another graphic interface (for example JSPUI). Examples of code are given in monospaced font.

Title, description, and keywords tags in a website header

This criterion requires filling in the specific metadata (e.g., metacontent) fields in the page's HTML code. The search engines process them automatically to find out what the website is about.

To solve these SEO issues change a website title (in default “DSpace Home”) located in the language translations config files in the folder path.

/dspace/webapps/xmlui/i18n/messages_en.xml.

Find the appropriate key and change the value. All content in this file is fully customizable.

Next, edit DSpace’s page structure config file (in path

/themes/Mirage/lib/xsl/core/page-structure.xsl)

in order to add the metadata content:

- a meta-description tag
- a keywords tag
- an author tag

with a carefully selected content and length just below the main <head> tag, as shown in the example:

```
<head>
<meta content="The Digital Repository receives, stores, indexes, preserves and disseminates
the digital content created by the Department of Media and Cultural Heritage"
name="description" />

<meta name="keywords" content="digital repository, The Department of Mediamatics and Cultural
heritage, KMKD, MKD, Faculty of humanities, digital archive, DSpace"/>

<meta content="Matúš Formanek" name="author" />
```

Note: Do not forget the termination characters />. The keywords should be included in title and meta-description tags.

Several other SEO parameters are affected by performing those steps, for example, Google Search Results Preview test, Keywords Usage test, Unique key words test and keywords concentration test.

Language declaration

The language declaration is very important for search engines to identify the primary language of the website content.

If a declared language is missing in a website, you can define it by adding the following line into the page-structure.xsl file (the process is similar to adding keywords and description tag as explained above). Edit the page-structure.xsl file (with VIM or another text editor, for example) and add a statement like the following above the main <head> tag:

```
<html xml:lang="sk" xmlns="http://www.w3.org/1999/xhtml"></html>
```

Note: “sk” is an abbreviation for “Slovak language” as stated in W3 namespaces. More information is available at <https://www.w3.org/TR/xml/>.

Google Analytics, robots.txt and sitemap implementation

The connection between a website and Google Analytics services enables Google Analytics to track users’ behavior and understand as they interact with this site. It is the basis of web analysis. The “robots.txt” and “sitemap.xml” files are simple text files which are required for search engines to specify the website structure and additional information about it.

To enable Google Analytics services, insert a UA code identifier (ID is a string), obtained from Google Analytics, into the DSpace.cfg config files located in the DSpace home folder. In that file find the key/row named “`xmlui.google.analytics.key=`” and insert the corresponding UA identifier there. Next, it is needed to uncomment the row with the key “`xmlui.controlpanel.activity.max = 250`” in the same “DSpace.cfg” file.

Finally, uncomment the row below in the “`xmlui.xconf`” file located in the path `/DSpace/config/` and restart the Tomcat service:

```
<aspect name="StatisticsGoogleAnalytics" path="resource://aspects/StatisticsGoogleAnalytics/" />
```

The “`robots.txt`” file is commonly used and enabled in DSpace, but many SEO audit tools are not able to detect it successfully because this file is located in path other than the expected default one. To enable `robots.txt` file detection, copy the file `/DSpace/webapps/xmlui/static/robots.txt` to the root of the Tomcat folder (usually located in path `/var/lib/tomcat8/webapps/ROOT`). Finally, restart the Tomcat web service.

A sitemap for a currently running DSpace instance is available in the “`robots.txt`” file mentioned above. Edit this file and set an appropriate URL for the sitemap location.

Enabling connections with social media

This criterion detects a hyperlink (or other metadata) connection between a website and popular social media, such as Facebook, Twitter, etc. The primary goal is to promote the digital content.

This subsection deals with social media connections with a DSpace-based repository. A simple creation of a profile or a site on a social network related to a repository is considered an essential example of good practice. However, an appropriate form of connection between sites must be created, too. Naturally, further endorsement of this system through social networks is another key step.

Social media-oriented tests are performed by each SEO audit tool nowadays. The detected connection with social media could have a big impact on the site’s popularity, as well as on the gained SEO final score. There are many ways how to establish these connections:

Connection with Facebook, Instagram and Twitter—a direct link from the homepage, for example: To add a link to a Facebook site profile, edit the page-structure file (`/DSpace/webapps/xmlui/themes/Mirage/lib/xsl/core/page-structure.xsl`) just below a div tag with ID “`ds-footer-wrapper`”. For example:

```
<div id="ds-footer-wrapper">
  <div id="ds-footer">
    <div id="ds-footer-left">
      <a href="https://www.facebook.com/digitalnyrepozitar/" target="_blank">Facebook page</a>
    </div>
```

A direct link to other media could be similarly added, too, if needed. After accomplishing this procedure, the test for social media (e.g., Facebook) passes correctly. However, it should be done by adding microdata or other inserted structured metadata into the html code. Please, do not forget to add the appropriate links in the schema.org metadata, too (see below). It is advisable to periodically promote the repository through the posts on Facebook and other social media.

Twitter optimization through a Twittercard insertion—Some audit tools perform a specific test focused on a presence of so-called Twittercard HTML markup on the website to optimize future Tweets. Only one card type per-page is supported. Here is an example for Twittercard consisting of four parts (card, title, image, description).

After customizing the following code to reflect your institution’s information, insert the following code just below the <head> tag in the page-structure.xml file:

```
<meta name="twitter:card" content="Digital repository of The Department of Mediamatics and cultural heritage"/>
<meta name="twitter:title" content=" Digital repository of The Department of Mediamatics and cultural heritage "/>
<meta name="twitter:image" content="https://opensenselabs.com/sites/default/files/inline-images/Screenshot%202018-07-02%2019.25.45.png"/>
<meta name="twitter:description" content=" The Digital Repository receives, stores, indexes, preserves and disseminates the digital content created by the Department of Media and Cultural Heritage "/>
```

Note: The URL of the twitter:image must be absolute.

Placing a repository link into other (especially high-ranked) websites is also highly recommended. The gained score as well as website traffic will surely rise by implementing those links.

OpenGraph protocol integration

This criterion refers to the presence of specific metadata object (OpenGraph element) in the website. It is very important for website objects’ visibility in social networks. Many website objects could be described through OpenGraph metadata protocol tags.²⁵

The WooRank audit tool verifies whether OpenGraph tags on your webpage have been detected or not. OpenGraph protocol adoption is a way to enable the integration of any website with social media or other platforms. You will be able to control how your websites are presented when they (e.g., their links) are shared across social media with metadata stored in OpenGraph protocol tags (all documentation is available at <https://ogp.me/>).

To adopt the main OpenGraph elements, insert the following code (updating it for your institution) just below the <head> tag in the page-structure.xml file:

```
<meta property="og:title" content="Digital repository of The Department of mediamatics and cultural heritage"/>
<meta property="og:url" content="https://repozitar.kmkd.uniza.sk/xmlui"/>
<meta property="og:description" content=" The Digital Repository receives, stores, indexes, preserves and disseminates the digital content created by the Department of Media and Cultural Heritage "/>
<meta property="og:locale" content="sk_SK"/>
<meta property="og:locale:alternate" content="en_US"/>
<meta property="og:type" content="website"/>
<meta property="og:image:type" content="text/html"/>
<meta property="og:image" content="https://opensenselabs.com/sites/default/files/inline-images/Screenshot%202018-07-02%2019.25.45.png"/>
```

Note: The URL of the tag “og:image” must be absolute.

Structured data integration (schema.org)

This criterion analogously deals with the presence of objects described by another standard for structured metadata, schema.org:

Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet. Schema.org uses vocabularies that can be used with many different encodings, including RDFa, Microdata and JSON-LD.²⁶

Google, Microsoft, and others already use these vocabularies to power rich, extensible experiences with a shared collection of schemas.²⁷ A lot of information included in a DSpace repository can be described by schema.org vocabulary. There are three main ways how to do that—through JSON-LD, RDFa, and Microdata.

JSON-LD is a JavaScript notation embedded in a <script> tag in the page head or body. The markup is not interleaved with the user-visible text, which makes nested data items easier to express, such as the Country of a PostalAddress of a MusicVenue of an Event. Also, Google can read JSON-LD data when it is dynamically injected into the page's contents.

RDFa is an HTML5 extension that supports linked data by introducing HTML tag attributes that correspond to the user-visible content that you want to describe for search engines. RDFa is commonly used in both the head and body sections of the HTML page.

Microdata is an open-community HTML specification used to nest structured data within HTML content. Like RDFa, it uses HTML tag attributes to name the properties you want to expose as structured data. It is typically used in the page body, but can be used in the head.²⁸

Google recommends using JSON-LD for structured data whenever possible.²⁹

We also recommend using JSON-LD (an extension of original JSON suitable for linking data) vocabulary to express additional information about entities related to a DSpace repository, for example information about the organization and many other entities.

To enter schema.org elements expressed in JSON-LD in order to increase the searching impact of a DSpace instance on the search engines, insert a script like following into the page-structure.xsl file (/DSpace/webapps/xmlui/themes/Mirage/lib/xsl/core/page-structure.xsl) just below a div tag marked with ID "ds-footer-wrapper".

```
<script type="application/ld+json">
{
  "@context" : "http://schema.org",
  "@type" : "EducationalOrganization",
  "name" : "Digital repository of The Department of mediamatics and cultural heritage",
  "department" : "The Department of mediamatics and cultural heritage",
  "url" : "https://repozitar.kmkd.uniza.sk/xmlui",
  "sameAs" : [
    "https://www.facebook.com/digitalnyrepozitar/",
    "https://www.instagram.com/katedramkd/",
    "https://twitter.com/MediamatikaKD"
  ],
  "address": {
    "@type": "PostalAddress",
    "streetAddress": "Univerzita 8215/1",
    "addressRegion": "Zilina",
    "postalCode": "01026",
    "addressCountry": "SK"
  }
}
</script>
```

Many elements like the one above could be added, too. See the documentation available at <https://schema.org/>. A free online tool for testing structured data is available from Google at <https://search.google.com/structured-data/testing-tool?hl=en>.

Reducing repository website size during transfer

The primary goal of this criterion is to measure a reduction of website size which is conducted by the enabled compression of website code parts.

This reduction can be achieved by enabling compression methods for HTML and other file formats when they are transmitted from a server to a client. The Tomcat webserver (which is an essential website component for DSpace repositories) allows turning on GZIP compression and so-called JavaScript minification.

To enable GZIP on the Tomcat webserver, edit the Tomcat's configuration file "server.xml" located in its home directory. Under the tag "<Service name="Catalina">" edit a corresponding connector tag so it looks like the following example. Changes in the code are shown in bold:

```
<Connector port="8080" maxHttpHeaderSize="8192"
  maxThreads="150" minSpareThreads="25" maxSpareThreads="75"
  enableLookups="false" redirectPort="8443" acceptCount="100"
  connectionTimeout="20000" disableUploadTimeout="true"
  compression="on"
  compressionMinSize="1024"
  noCompressionUserAgents="gozilla, traviata"
  compressableMimeType="text/html, text/xml, text/plain, text/javascript, text/css, application/
x-javascript, application/javascript"/>
```

The CompressableMimeType contains the formats you want to compress.

Important note: If you deal with HTTPS (and corresponding port number 443 instead 8080), you must set the options stated above into the corresponding connector (443), too. Otherwise, the compression will be enabled only in simple HTTP (running on port number 8080).

JavaScript minification can be enabled in a "DSpace.cfg" configuration file located usually in a DSpace home directory (/DSpace/config/). Change the key value from false to true in the following rows:

```
xmlui.theme.enableMinification = true
xmlui.theme.enableConcatenation = true
```

Setting a canonical link

This requirement deals with the presence of a canonical link used by search engines. "A canonical link is included in the HTML code of a webpage to indicate the original source of content. This markup is used to address SEO problems with duplicate content which arise when different pages with different URLs contain identical or nearly identical content."³⁰ The problem with duplicated content can arise, for example, when a webpage is accessible with or without a www prefix in its URL or a webpage is accessible via HTTP and HTTPS protocols. "For SEO purposes, the canonical link shows Google and other search engines which URL corresponds to the original source of content and should be shown in search results. It is added as a meta tag to every URL version of a given webpage and indicates the canonical URL."³¹

After a necessary customization, insert the following row just below the <head> tag in DSpace's page-structure.xml file

```
(/DSpace/webapps/xmlui/themes/Mirage/lib/xsl/core/page-structure.xml) :
```

```
<link rel="canonical" href="https://repozitar.kmkd.uniza.sk/xmlui/" />
```

HTTPS adoption

The adoption of HTTPS is required for a secure data transfer. This criterion inspects if HTTPS is enabled and what quality it displays.

HTTPS is an essential component that supports website security for sites available via the Internet. We pointed out the importance of HTTP adoption in a DSpace respiratory interface in our previous research papers.³²

Firstly, you should prepare a file called the Certificate Signing Request, or CSR, that will be used by the Certificate Authority of your choice to generate the Certificate SSL. The process of HTTPS configuration on the Tomcat webserver (used natively in DSpace repositories) is widely described online (for example available at <https://www.mulesoft.com/tcat/tomcat-ssl>). Secondly, you should configure a corresponding connector for HTTPS (port 443) in Tomcat's configuration file. We strongly recommend following those instructions and to use DSpace instance only with HTTPS, among other major security risks, because dealing with simple HTTP has surely a very negative impact on SEO final score. Google and other search engines strongly prefer websites with HTTPS enabled.

Discussion about the SEO Issues Solving Process

In the previous subsections, we have offered solutions of selected major SEO issues that can be relatively easily resolved in systems based on DSpace and its website technologies. However, in practice, it is unrealistic to expect a 100% optimization level and final solutions for all detected problems. Therefore, we intentionally did not mark the second state of the system (shown in table 1) as fully optimized but only semi-optimized. Some of the issues we detected remain unsolved despite all our efforts. There are several reasons. One of the most important of them is the fact that DSpace software, like many complex systems, cannot be easily modified without programming experience. Therefore, resolving some complicated issues is beyond the scope of this article. Another significant reason is that we lacked knowledge about some issues at the time of writing this paper and therefore we could not solve them. This situation creates an opportunity for further research and proposals for solutions of unsolved issues in this specific area, which the professional public would certainly like to welcome.

Taken together, it could be said that the changes we have made, helped to objectively increase the average SEO score by 59 percent compared to the default installation. All the successfully performed actions improved the search results of our repository and rapidly increased its.

We suppose that all related SEO actions can affect website traffic. Most major issues discussed in this case study were resolved before November 10, 2019. Therefore, we prepared an analysis of the repository traffic which involved 30-day period before and after this date (one from October 11 until the change, the other from the change until December 10, 2019). We determined the impact of performed SEO actions on website traffic. The results are satisfactory because the number of established relations has significantly raised. The impact of organic search (through Google, for example) has increased traffic by 47.67% (from 86 to 127 sessions). The number of so-

called referral sessions (sessions initiated from social media and other referral sites) has increased by 193.75% (from 16 to 47 sessions). Users spent much more time on the website and viewed more pages on average (an increase of up to 159%).

We view the significant traffic increase as a proof that the SEO changes we implemented helped to promote use of the digital repository's content.

In the next section, we want to compare the quantitative improvement of SEO parameters, which we have been able to achieve to this point, with the results achieved in global testing of worldwide DSpace-based repositories by the same set of tests. Next, we can easily compare the results gained in the local case study with the current state determined in the worldwide area of DSpace repositories.

TESTING SEO PARAMETERS OF WORLDWIDE DSPACE-BASED REPOSITORIES

There are several thousand digital repositories around the world. Most of them (41.1% according to ROAR registry data and over 39% according to the OpenDOAR registry) are based on DSpace software.³³ Therefore, we also focus our research exclusively on DSpace-based repositories in this study.

As we have pointed out in the methodology, the second objective of this paper is to briefly describe a current state of SEO parameters related to worldwide DSpace-based digital repositories. Next, we will discuss the comparison of results obtained from the case study and exploration of worldwide repositories.

Methodology

According to the facts stated above, we would like to know more details about the quality of SEO parameters related to worldwide repositories running with DSpace. We decided to use one of the two most authoritative registries of digital repositories: the Registry of Open Access Repositories (ROAR) and the Directory of Open Access Repositories (OpenDOAR). ROAR is hosted at the University of Southampton in the United Kingdom and is available online at <http://roar.eprints.org/>. OpenDOAR is available at <https://v2.sherpa.ac.uk/opensoar/>. Both are quality-assured global directories of academic open access repositories. They “enable the identification, browsing and search for repositories, based on a range of features, such as location, software or type of material held.”³⁴ We decided to utilize the ROAR registry as the source for a sample list because it is possible to filter systems based on specific criteria. We applied these three filters on March 11, 2020: any country, any repository type, and DSpace software. We downloaded the raw data in a text/CSV file with 1,977 records. Each record had a separate row for each repository. Each row has a sequence number and includes many columns with much additional information. Only a few columns were necessary for our purpose—the columns marked as “title” and “home_page”. Other columns were removed. All changes in the list were performed using Microsoft Excel.

For further evaluation, we selected a random sample from this file. We used a sample size online calculator (available at <https://www.calculator.net/sample-size-calculator.html>) to do that. We had set the following values for statistical parameters:

- Population size: 1,977 (the total count of DSpace repositories in ROAR)
- Confidence level: 95%
- Margin of error: 10%

A sample size of 92 was automatically calculated for these values of statistical parameters. Next, we used a random number generating function integrated in Excel (randbetween(1,1977)) that generated 92 random numbers from the strictly defined range. Each randomly generated number corresponds with the matching row number in the table of repositories downloaded from the ROAR. We could choose 92 DSpace repositories for testing purposes. In this way, objectivity in the selection of the research sample was ensured.

We also tested the sample for duplicate entries, to ensure that no repository was selected twice. We had to do so, because the random generating function does not guarantee that only unique integer values will be generated. Figure 1 shows the distribution histogram of randomly generated values from 1 to 1,977.

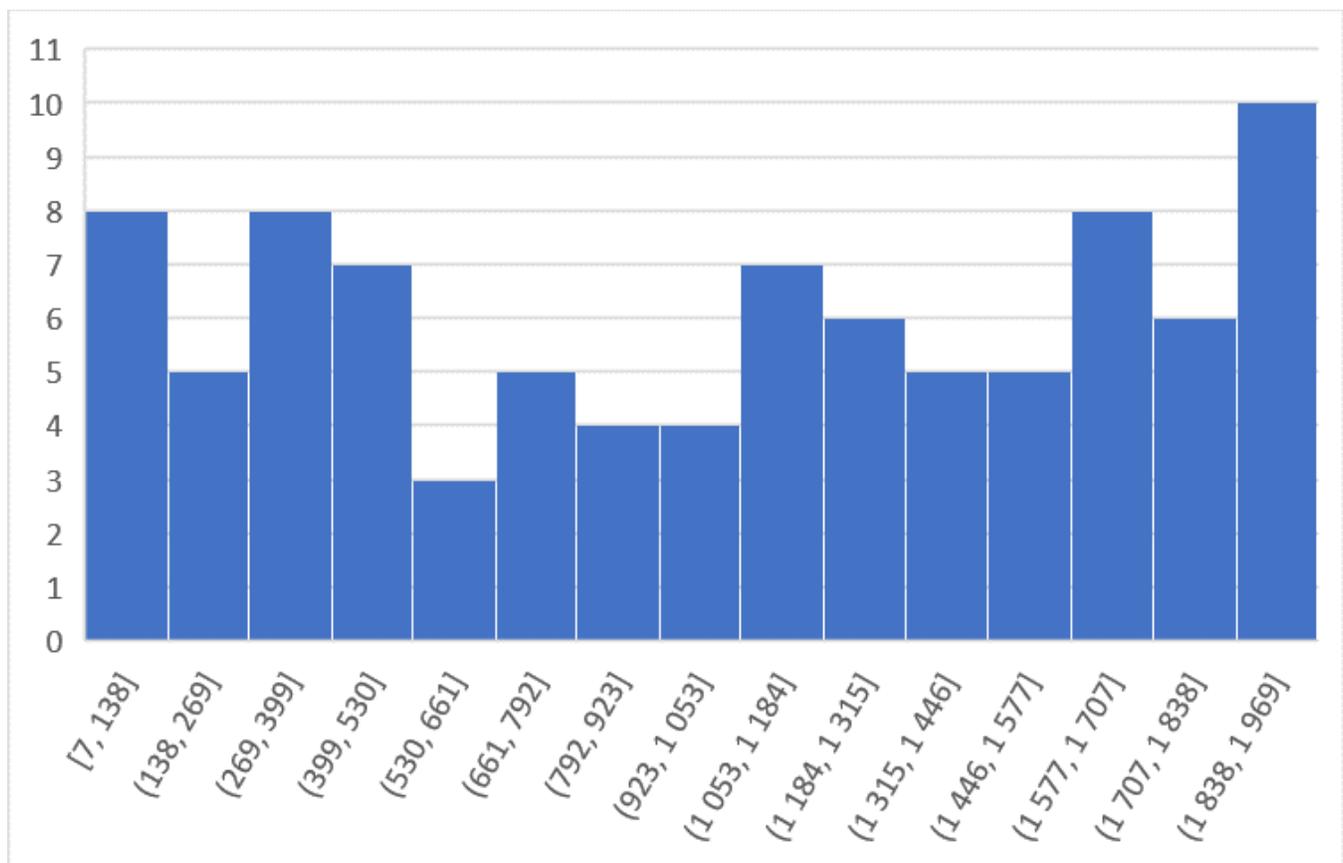


Figure 1. Distribution histogram of randomly generated values.

Then, we attempted to test each of 92 selected repositories with three audit tools. The results are discussed in the next section.

Test Results

Table 2 shows a part of the table with results. This table does not contain any URLs or titles to ensure anonymity, however we can provide this information upon request. A second-level domain name is only displayed in each row as well as the corresponding scores gained in tests. The maximum value is 100 points in every case. The rows in the table were sorted by the calculated average score from high to low. Many rows are omitted due to the table length (one row for the each from 92 repositories). The last tested repository has a sequence number equal to 65. The repositories with a higher sequence number have no gained score (N/A state), due to inaccessibility.

Table 2. Test results

The repository sequence number	The first and second-level domain name	SEO Site Checkup	SEO Checker	WooRank	Average
1	econstor.eu	76	65.9	69	70.30
2	datadryad.org	73	54.9	54	60.63
3	edu.ar	66	54.5	61	60.50
4	cuni.cz	60	52.8	65	59.27
5	edu.co	65	55.5	56	58.83
...
65	ac.cn	36	21.7	33	33.23
66	Scholarportal.info	N/A	N/A	N/A	N/A
...
89	org.br	N/A	N/A	N/A	N/A
90	mapfig.com	N/A	N/A	N/A	N/A
91	edu.ec	N/A	N/A	N/A	N/A
92	edu.co	N/A	N/A	N/A	N/A
Average score gained from particular tests		53.47	48.08	49.22	50.26
Standard deviation		9.31	9.29	10.27	9.62
Median		54	46.7	52	50.90
Modus		52	40	54	48.67

The testing process started on March 11, 2020, and finished on April 6, 2020. It took a lot of time, because we were limited by the reuse restrictions (described above) in the audit tools' free accounts. These restrictions meant that only a few tests could be performed daily even though we used several public IP addresses to speed up the overall testing process.

Among other items, we identified a startling problem related to nonfunctional repository URLs. Thirty one out of 92 tested repositories were unavailable between March and April 2020 (in table 2, they are shown with N/A status). On April 6, 2020, at the end of testing period, we tried to test the unavailable systems once again. Four of them had become functional, so the final number of

really tested repositories rose to 65 (out of 92). The remaining 27 (29.35 percent of the total) repositories were still offline or unavailable. It is possible that the URLs stated in ROAR’s records have been out-of-date. N/A values were ignored in all calculations and had no impact on the final average score or other statistical parameters. Only 65 fully functional DSpace-based worldwide repositories were involved and were used for testing purposes.

For better visualization of the partial as well as summarized results, we have decided to use a graph instead of a table. Figure 2 shows the results of 65 repositories sorted by an average gained score (from highest to lowest) that was calculated from three partial scores gained in SEO Site Checkup, SEO Checker and WooRank testing tools. So, there are three corresponding partial discrete values (colored dots) shown for each repository in figure 2. The calculated average score for each one is marked in red color. The red dotted line provided the most valuable results for this partial section.

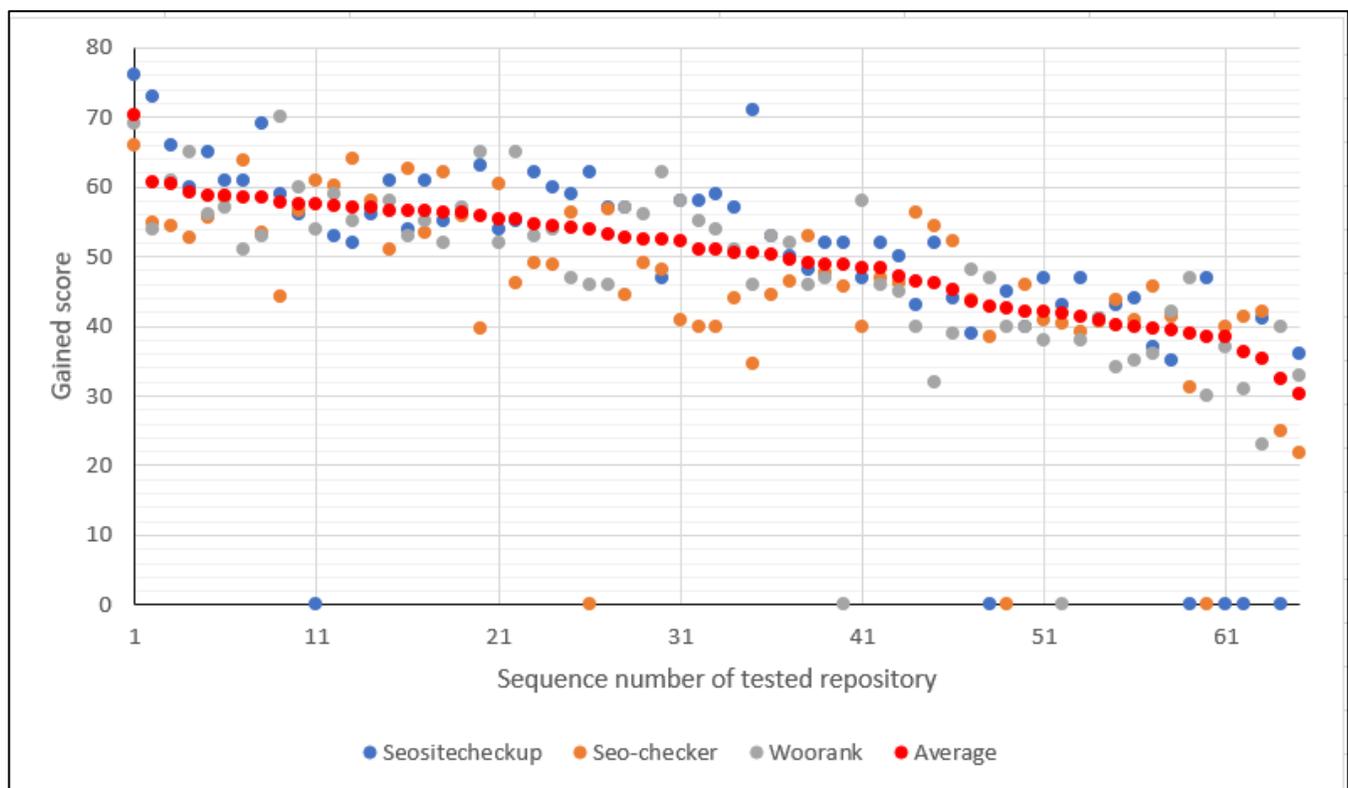


Figure 2. Results of 65 involved repositories.

The repositories that gained a higher score (e.g., better SEO results) are, logically, situated on the left side of figure 2. On the right side are systems with lower scores. Non-functional systems (N/A) are not displayed at all.

The underlying frequency distribution graph of average score (the red dotted line in the previous figure) is available in figure 3.

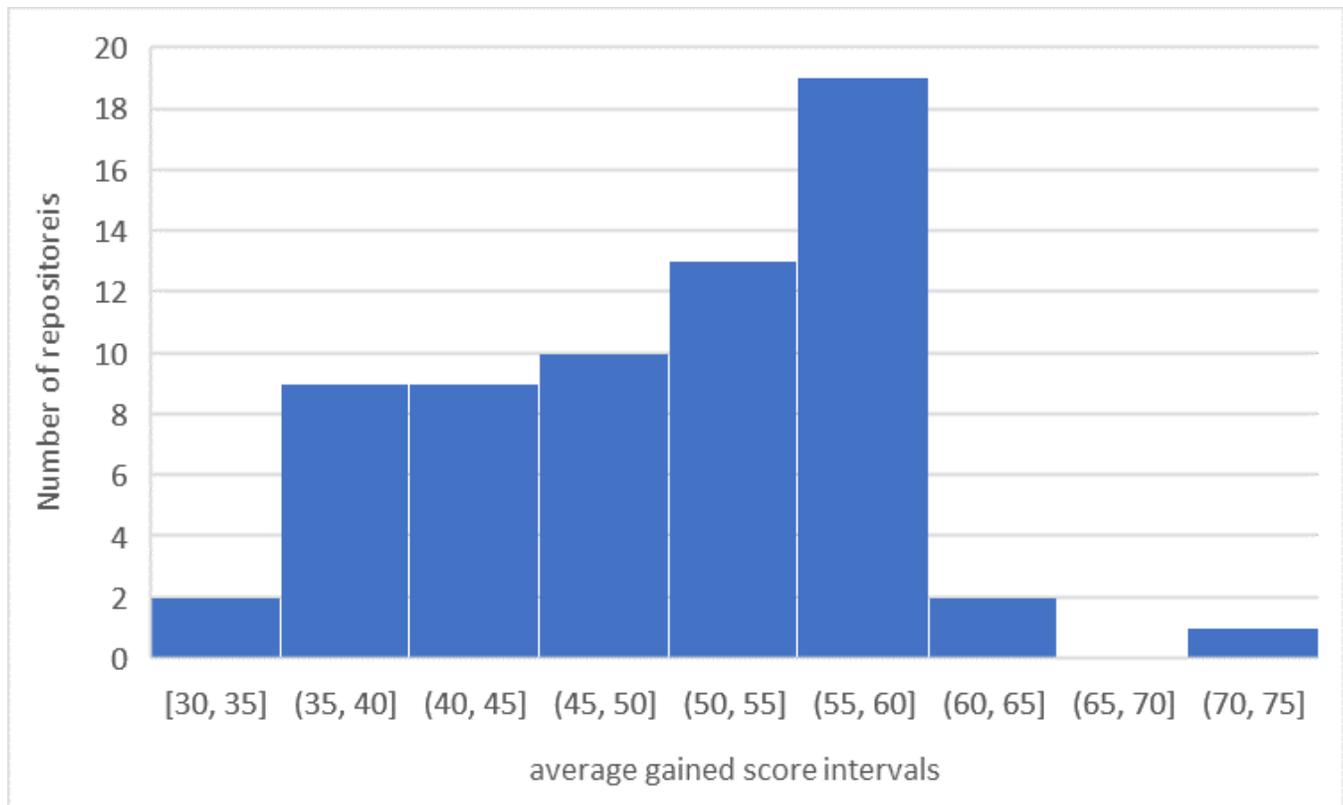


Figure 3. Underlying frequency distribution graph of average score

Based on the submitted results manifested in the previous figures we can make the conclusion with a relatively high degree of reliability:

The large part of DSpace-based repositories registered in ROAR (**over 29%**) **were unavailable** at the time of writing the article. It is alarming, because ROAR is still considered as an authoritative registry for open access repositories and should not contain any invalid data.

An average score of functional repositories, gained during the testing period, is very similar between audit tools: 53.47 points in SEO Site Checkup, 48.08 points in SEO Checker and 49.22 points in WooRank. Standard deviations of population are comparable, too. Finally, most of the tested repositories (19) gained a score from the interval (55,60) as is shown in figure 3; however, the average SEO score of all tested DSpace-based repositories **was only 50.26 points out of 100** (data from March/April 2020), which is an adequate value for a relatively low level of search engine optimization of those systems.

Results and Discussion

We have obtained complete insights on the SEO parameters of worldwide DSpace-based digital repositories in the previous section. Now, we can compare this data with the results gained during the case study solving process described above. The situation is briefly pointed out in table 3.

Table 3. Comparison of fresh installation, semi-optimized installation and average worldwide score

	SEOSitecheckup	SEO checker	Woorank	Total average score	Calculated improvement (%)
Fresh DSpace installation	58	50.1	32	46.7	100 - a reference point
Semi-optimized state of the Institutional Repository of The Department of Mediamatics and Cultural Heritage	81	78	59	74.66	+59.87
The average score of worldwide DSpace-based repositories	53.47	48.08	49.22	50.26	+7.62

Based on table 3, it is proposed that the fresh, non-optimized DSpace obtained a slightly worse score than the worldwide average. Although a few SEO issues still remain in our semi-optimized DSpace instance, the state of SEO parameters is much better than the score gained in any other discussed cases. If we considered a fresh DSpace installation as a reference point (100 percent), the improvement level would be shown in the last column of table 3. Semi-optimized DSpace offers an improvement up to 59.87% compared to fresh (non-optimized) DSpace installation. There is no significant difference (up to 7.62%) in SEO quality between the worldwide average repository and non-optimized DSpace instance. The results they have obtained are very similar.

As we have mentioned at the beginning of our paper, a higher score obtained in tests is not the primary objective. The main goal is to improve visibility and the content searchability of digital repositories, as well as to improve their security and ways of promotion through the social/new media.

CONCLUSION

This study exposed a serious research in the field of digital repositories running DSpace software—as the most popular tool for this purpose. We have shown that significant SEO improvement of more than 59% can be achieved thanks to a few simple modifications within the DSpace configuration and associated used application layers (Tomcat webserver, etc.). Some of those technical optimization steps can be performed in a relatively simple way, using previously described solving procedures and a wide theoretical background. We have publicly presented the reports and solving explanations of the most common and major SEO problems that DSpace repositories usually face. This paper is one of the first academic studies to deal with SEO issues related to digital repositories, especially those that are running DSpace software. We realize that we have not been able to solve all of the identified problems completely. Therefore, the following SEO issues remain unresolved:

- H1 Heading Tags test
- H1 Coherence test
- SEO friendly URL test
- Inline CSS test
- Page Cache test (Server-side Caching)
- CDN Usage test
- Image, Javascript, CSS Caching tests
- CSS Minification test
- URL Canonicalization test

Some of these could probably be solved more easily than others; however, the system URL cannot be changed without difficulty to be considered as SEO friendly.

In conclusion, all of the above presents a great opportunity for further discussions and research in this field. The current state of SEO parameters related to DSpace repositories has been presented as unsatisfactory, as shown in the test results. Conclusively, the results of our research indicate that there is a small difference in SEO quality between the average results obtained by global, worldwide DSpace repositories and the non-optimized installation of DSpace v6.3 (the difference is approximately 7% in global repositories' favor). It seems that the most of these systems are not currently optimized in terms of SEO and other technical website parameters. The second major finding indicates that the metadata records stored in the ROAR are not always accurate and may be incorrect or obsolete. In order to make this finding more objective we must note that the ROAR's storage had a major failure, which could lead to the harvesting service failing. (More information about the failure is available at [http://roar.eprints.org/.](http://roar.eprints.org/))

Finally, we recommend periodically re-testing the level of search engine optimization on digital repositories. The "search engine algorithms tend to change often, and new factors are added while outdated or not effective factors are excluded. This is why web developers must check the algorithm changes and adjust their websites in order to not only achieve but also maintain high ranking in search engines."³⁵

We believe that our work will also contribute to the initiation of cooperation among other experts in order to resolve remaining SEO problems. Ultimately, we hope that all presented efforts and recommendations will help repository administrators, users, scientists, researchers, teachers as well as students and other members of the general public to find what they need in the virtual spaces like digital repositories more quickly and efficiently.

ENDNOTES

¹ Christos Ziakis et al., "Important Factors for Improving Google Search Rank," *Future Internet* 11, no. 2 (January 2019): 2–3, <https://doi.org/10.3390/fi11020032>.

² F. Insidro Aguillo et al., "Comparing university rankings," *Scientometrics* 85 (February 2010): 243–56, <https://doi.org/10.1007/s11192-010-0190-z>.

³ Ahmad Bakeri Abu Baka and Nur Leyni, "Webometric Study of World Class Universities Websites," *Qualitative and Quantitative Methods in Libraries* (July 2017): 105–15, <http://qqml-journal.net/index.php/qqml/article/view/367>; Andreas Giannakoulopoulos et al., "Academic

- Excellence, Website Quality, SEO Performance: Is there a Correlation?" *Future Internet* 11, no. 11 (November 2019): 242, <https://doi.org/10.3390/fi11110242>.
- ⁴ Dwi Budi Santoso, "Pemanfaatan Teknologi Search Engine Optimazion sebagai Media untuk Meningkatkan Popularitas Blog WordPress," *Dinamik* 14, no. 2 (2009): 12–33, <https://www.unisbank.ac.id/ojs/index.php/fti1/article/view/100>; M. Iskandar and D. Komara, "Application Marketing Strategy Search Engine Optimization (SEO)," IOP Conference Series: *Materials Science and Engineering* 407 (2018), <https://iopscience.iop.org/article/10.1088/1757-899X/407/1/012011/pdf>.
- ⁵ Giannakouloupoulos et al., "Academic Excellence, Website Quality, SEO Performance."
- ⁶ Giannakouloupoulos et al., "Academic Excellence, Website Quality, SEO Performance."
- ⁷ Thomas Abrahamson, "Life and Death on the Internet: To Web or Not to Web is No Longer a Question," *Journal of College Admission* 168 (2000): 6–11.
- ⁸ Sukhpuneet Kaur, Kulwant Kaur, and Parminder Kaur, "An Empirical Performance Evaluation of Universities Website," *International Journal of Computer Applications* 146, no. 15 (July 2016): 10–16, <https://doi.org/10.5120/ijca2016910922>.
- ⁹ "Best SEO Software," G2, last modified 2020, <https://www.g2.com/categories/seo>.
- ¹⁰ "Best SEO Software."
- ¹¹ Ziakis et al., "Important Factors for Improving Google Search Rank."
- ¹² Giannakouloupoulos et al., "Academic Excellence, Website Quality, SEO Performance."
- ¹³ Joeran Beel, Bela Gipp, and Eric Wilde, "Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature for Google Scholar & Co.," *Journal of Scholarly Publishing* 41, no. 2 (January 2010): 176–90, <http://dx.doi.org/10.3138/jsp.41.2.176>.
- ¹⁴ Brian Kelly, "MajesticSEO Analysis of Russell Group University Repositories," *UK Web Focus* (blog), August 29, 2012, <http://ukwebfocus.wordpress.com/2012/08/29/majesticseo-analysis-of-russell-group-university-repositories/>.
- ¹⁵ "OpenDOAR Statistics," Jisc, last modified September 2020, https://v2.sherpa.ac.uk/view/repository_visualisations/1.html.
- ¹⁶ Si Ong Quan, "44 Best Free SEO Tools (Tried & Tested)," last modified May 28, 2020, <https://ahrefs.com/blog/free-seo-tools/>; Navneet Kaushal, "Top 15 Most Recommended SEO Tools," last modified September 2020, <https://www.pagetraffic.com/blog/top-15-most-recommended-seo-tools/>.
- ¹⁷ Quan, "44 Best Free SEO Tools (Tried & Tested)"; Kaushal, "Top 15 Most Recommended SEO Tools."

-
- ¹⁸ “28 Top SEO Site Checkup Tools,” Traffic Radius, accessed March 29, 2020, <https://trafficradius.com.au/seo-site-checkup-tools/>.
- ¹⁹ “28 Top SEO Site Checkup Tools,” Traffic Radius.
- ²⁰ Chandan Kumar, “13 Online Tools to Analyse Website SEO for Better Search Ranking,” last modified April 11, 2020, <https://geekflare.com/online-tool-to-analyze-seo/#SEO-Tester-Online>.
- ²¹ “28 Top SEO Site Checkup Tools,” Traffic Radius.
- ²² Kumar, “13 Online Tools to Analyse Website SEO for Better Search Ranking.”
- ²³ “OpenDOAR Statistics,” Jisc.
- ²⁴ “DSpace 7.0 Beta 4 Release Announcement,” Lyrasis, October 13, 2020, <https://duraspace.org/dspace-7-0-beta-4-release-announcement/>.
- ²⁵ “The Open Graph protocol,” OGP, accessed January 25, 2021, <https://ogp.me/>.
- ²⁶ “Welcome to Schema.org,” Schema, accessed May 1, 2020, <https://schema.org/>.
- ²⁷ “Welcome to Schema.org.”
- ²⁸ “Understand how structured data works,” Google, accessed May 2, 2020, <https://developers.google.com/search/docs/guides/intro-structured-data>.
- ²⁹ “Understand how structured data works,” Google.
- ³⁰ “Canonical Tag,” Seobility, accessed March 20, 2020, https://www.seobility.net/en/wiki/Canonical_Tag.
- ³¹ “Canonical Tag,” Seobility.
- ³² Matus Formanek and Martin Zaborsky, “Web Interface Security Vulnerabilities of European Academic Repositories,” *LIBER Quarterly* 27, no. 1 (February 2017): 45–57, <http://doi.org/10.18352/lq.10178>; Matus Formanek, Vladimir Filip, and Erika Sustekova, “The Progress of Web Security Level Related to European Open Access LIS Repositories between 2016 and 2018,” *JLIS.it* 10, no. 2 (May 2019): 107–15, <http://dx.doi.org/10.4403/jlis.it-12545>.
- ³³ “OpenDOAR Statistics,” Jisc.
- ³⁴ “About OpenDOAR,” Jisc, last modified September 2020, <https://www.jisc.ac.uk/opendoar>.
- ³⁵ Ziakis et al., “Important Factors for Improving Google Search Rank,” 2.