

Article

Big Data Analytics for Search Engine Optimization

Ioannis C. Drivas ^{1,*} , Damianos P. Sakas ², Georgios A. Giannakopoulos ¹ and Daphne Kyriaki-Manessi ¹ 

¹ Department of Archival, Library and Information Studies, Lab of Information Management, University of West Attica, Ag. Spyridonos, 12243 Egaleo, Greece; gian@uniwa.gr (G.A.G.); dkmanessi@uniwa.gr (D.K.-M.)

² School of Applied Economics and Social Sciences, Agricultural University of Athens, Iera Odos 75, 11855 Athens, Greece; d.sakas@aua.gr

* Correspondence: idrivas@uniwa.gr; Tel.: +30-69-7401-6823

Received: 8 March 2020; Accepted: 30 March 2020; Published: 2 April 2020



Abstract: In the Big Data era, search engine optimization deals with the encapsulation of datasets that are related to website performance in terms of architecture, content curation, and user behavior, with the purpose to convert them into actionable insights and improve visibility and findability on the Web. In this respect, big data analytics expands the opportunities for developing new methodological frameworks that are composed of valid, reliable, and consistent analytics that are practically useful to develop well-informed strategies for organic traffic optimization. In this paper, a novel methodology is implemented in order to increase organic search engine visits based on the impact of multiple SEO factors. In order to achieve this purpose, the authors examined 171 cultural heritage websites and their retrieved data analytics about their performance and user experience inside them. Massive amounts of Web-based collections are included and presented by cultural heritage organizations through their websites. Subsequently, users interact with these collections, producing behavioral analytics in a variety of different data types that come from multiple devices, with high velocity, in large volumes. Nevertheless, prior research efforts indicate that these massive cultural collections are difficult to browse while expressing low visibility and findability in the semantic Web era. Against this backdrop, this paper proposes the computational development of a search engine optimization (SEO) strategy that utilizes the generated big cultural data analytics and improves the visibility of cultural heritage websites. One step further, the statistical results of the study are integrated into a predictive model that is composed of two stages. First, a fuzzy cognitive mapping process is generated as an aggregated macro-level descriptive model. Secondly, a micro-level data-driven agent-based model follows up. The purpose of the model is to predict the most effective combinations of factors that achieve enhanced visibility and organic traffic on cultural heritage organizations' websites. To this end, the study contributes to the knowledge expansion of researchers and practitioners in the big cultural analytics sector with the purpose to implement potential strategies for greater visibility and findability of cultural collections on the Web.

Keywords: cultural analytics; cultural data; search engine optimization; SEO strategy; SEO factors; big data; websites visibility; predictive modeling; website security; website load speed; user behavior

1. Introduction

Over the last 10 years, big data analytics has been called “the oil” for optimizing the digital ecosystem and, subsequently, the World Wide Web sphere. Untamed big data with enormous volume and velocity are generated regarding the interactions of users with search engines and websites and how they respectively react to the search results and the content they receive. Search engines, search results, and websites express a cause-and-effect relationship under their main purpose of existence.

This purpose is related to the provision of the highest volume of information, in the best time, with the most precise results, according to the users' search terms.

In the context of the Big Data era, search engine optimization (SEO) plays a crucial role in the potential dissemination of personalized content that reflects quality. This quality is related to the curation of content and proper usability in the Web-based systems in order to cover users' information needs. SEO's main purpose is related to the provision of strategic steps to Web developers and content creators to optimize websites for higher rankings in search results and, hence, greater organic traffic that comes from search engines.

Nevertheless, despite the large volume of datasets and analytics that are produced and related to website performance and users' behavior inside them, the research field for utilizing big data analytics for strategic SEO schemas still remains in infancy. Several reasons are related with this research gap, such as symptoms of data overloading and, hence, difficulties in preprocessing, analysis, visualization, and interpretation of the outcomes. Subsequently, another drawback is related with the reduced managerial capabilities to understand the intercorrelations between the semantic Web analytics metrics and how they are aligned to SEO purposes. Another difficult point is the absence of methodological mechanisms that articulate validity, reliability, and consistency regarding the variables that are taken into consideration, with the purpose to optimize visibility of websites. Against this backdrop, this paper presents a novel methodological approach for utilizing big data analytics related to website performance and how they contribute to the SEO goal, which is an increase in organic search engine traffic percentage. We implement this methodology in the cultural heritage domain.

Cultural heritage institutions (CHIs) and their content in both the physical and digital worlds represent the social desire to preserve and cross-link the legacy of former generations in today's world, reflecting the ancestral identity of each society. Cultural institutions such as museums, galleries, libraries, and archives support the democratization of cultural heritage. They shape and reinforce the cultural and educational background of the people, while their mission is to combine the fragmented pieces of history depending on the society they belong to and serve. The information communication technologies and the Web expand the opportunities for cultural institutions to attract more stakeholders, both in physical and in online environments. This fact strengthens their scope for providing cultural knowledge to interested parties, global, national, and local communities, educating them in a quite attractive way through websites [1]. This involves a shift from traditional approaches of *keep and protect* to *experience and engage* in the era of digital humanities [2–4].

From a managerial perspective, the purpose of websites is to improve awareness, increase interest about artifacts, and holistically enhance the online visibility and findability of cultural collections. However, the nature of cultural heritage websites means that they deal with massive amounts of datasets, such as a high volume of internal webpages, links, images, and depth in user exploration and experience. This raises difficulties for managers to handle large-scale collections, increasing uncertainty about the level of visibility that cultural websites have on the Web. In this respect, it is a necessity to propose a SEO framework that utilizes generated big data analytics about CHI websites and their performance. In this way, the decision-makers in CHIs will benefit from a well-informed toolbox supportive of efforts to increase cultural content visibility on the Web.

To this end, this study unfolds in four parts. Firstly, we describe in a detailed manner the SEO factors that we assume impact the organic search engine visit percentage. After that, the methodology unfolds in three different stages. First, we summarize and validate the extracted big data analytics. Subsequently, we proceed into the development of diagnostic exploratory models that estimate the cause-and-effect relationships between the proposed factors. In the third stage of the methodology, we describe the initial steps for predictive model development and simulation for optimizing organic search engine traffic. The paper continues with the presentation and interpretation of results. Lastly, the discussion and conclusions are presented, suggesting practical managerial implications for the optimization of CHI websites in terms of performance, visibility, and findability on the Web.

2. The Nature of Cultural Heritage Websites

According to Wilson [5], in every cultural heritage institution (CHI), the first main purpose is the management of the cultural material they contain. The second purpose is to make that material accessible to everyone who wants to see it. Indeed, even from prior studies at the infancy phase of the Web, up to recent approaches, CHIs have utilized the opportunities of Web presence via websites in order to expand the visibility and accessibility of their cultural heritage content [2,3,6–8]. Websites under the paternity of CHIs constitute a vital online channel that improves awareness and expands capacities for visitation [7]. At the same time, it gives the advantage of presenting cultural information that is able to exceed by far that available at the physical place. Subsequently, research efforts by Voorbij [9] and Marty [10] demonstrated that the user experience across websites leads to better-prepared visitors and also to enhanced post-visit experience in terms of the meaning and interpretation they attribute to cultural collections.

However, the utilization of websites for CHIs with the purpose to optimize the visibility of their artifacts on the Web is not a straightforward and easy road. The CHI websites are constituted by massive amounts of datasets related to architectural aspects, such as the high number of webpages they contain, hundreds of images, thousands of internal and external interlinks, and so on. In this respect, prior efforts noted that the growth trend of digitizing artifacts is not sufficient or organized for the accurate retrieval of cultural information in large-scale Web-based collections. More specifically, this challenges visitors to face (a) overloading symptoms of the provided cultural heritage information [11] and/or (b) lack of cultural information due to low level of findability and visibility [12,13].

The two issues referred to above reflect the scarcity of an effective SEO approach that could be measured in a quantifying way to increase the accuracy of information retrieval from CHI websites. This statement has also been supported by a recent report by Krstic and Maslikovic. [2]. They showed that only one-third of their examined cultural institutions confirmed the utilization of analytics platforms for the evaluation and optimization of user experience during content search and navigation through digital collections. This kind of evaluation, intended to reinforce content that suffers from low levels of findability and visibility, becomes even more complex as CHI websites are characterized by the massive sizes of their content.

2.1. The Size of Cultural Heritage Websites

The size of cultural heritage websites can be set under the context of the Vs of big data. They contain a massive volume of cultural information integrated into unique webpages, images, and hyperlinks. At the same time, a huge number of visitors engage with the content, generating almost real-time behavioral datasets with high velocity. In terms of variety, different kinds of datasets from different smart devices are produced. This kind of data variety is related to web analytics metrics of interaction with the cultural content. Other kinds are related to unstructured datasets, such as images uploaded on social networks and text from reviews, as regards visitor opinion about cultural artifacts and their online representation [14]. Regarding the value of the produced big cultural analytics data, this is mostly related to the key point indicators that cultural institutions determine for their growth over time. For instance, the behavioral datasets generated from visitor engagement with the e-shop of the museum indicate economic value from data utilization. Similarly, the utilization of the generated datasets regarding the performance of CHI websites and their level of visibility in search engines point out societal and educational value. That is, the higher the visibility and findability, the greater the possibilities for passing on cultural heritage information to stakeholders for the improvement of their educational level.

Regardless of the type of organization, the larger it is, the greater the possibilities to utilize web data analytics strategies for potential evaluation and enhancement of its online presence [15]. However, big data analytics bring big troubles. It should be noted that the existence of a website with massive content and, thus, high volume of data production does not necessarily mean increased availability of suggestions and solutions for strategic SEO planning.

In the Big Data era, the latter assumption signals a contradictory element against prior efforts stating that “content is king” [16,17]. Controversially, the bigger the content, the harder the manipulation and the identification of technical and behavioral factors that probably influence the website’s performance in terms of its visibility and findability in search engine result pages. In fact, it is practically more manageable to optimize a website that contains up to 100 webpages, rather than a website with thousands of webpages, images, and, hence, complexity in link architecture. Preliminary findings of our research indicate that CHI websites are characterized by massive size in terms of webpages, links, and images (Table 1 and Figures 1–4).

However, although prior studies have shown that the implementation of SEO strategies returns higher rankings and possibly better organic traffic [16,18,19], it remains unclear whether the size of the website affects the organic search traffic percentage, that is, the bigger the content of the website in terms of webpages, images, internal, and external links, the greater the percentage of organic search engine traffic. This implication led us to develop the first hypothesis:

Hypothesis (H1). *The size of cultural heritage institutional websites impacts the percentage increase of organic search engine traffic that they receive.*

Table 1. Descriptive statistics from the examined 171 unique cultural heritage institution websites.

Descriptives per Unique Domain	Number of Webpages	Number of Images	Number of Internal Links	Number of External Links
Mean	835.77	509.74	1502.61	577.5
Min	36	47	163	29
Max	2088	1368	2716	2360
Std Dev	390.53	302.96	542.49	461.4

N = 171.

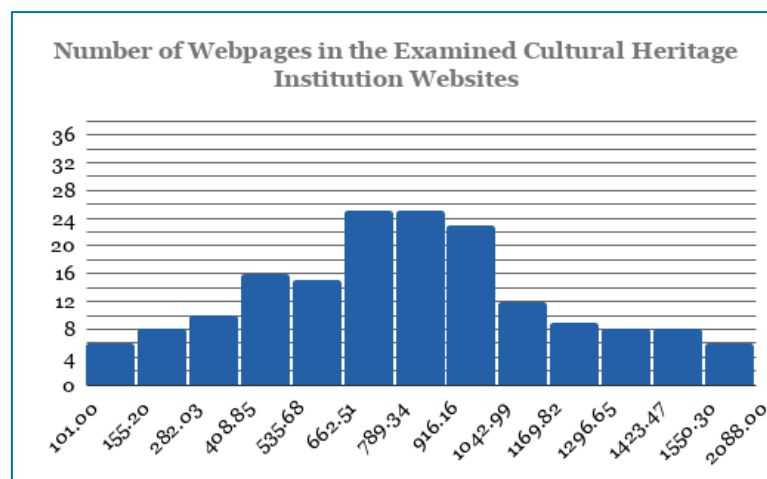


Figure 1. Histogram of the number of unique webpages that the examined CHI websites contain. The horizontal axis indicates the numerical amount of unique webpages that the examined websites contain. The vertical axis demonstrates how many websites contain from 101 up to 2088 unique webpages. For instance, there are 16 websites of the total 171 that ranged from 408 up to 535 unique webpages and 25 of the total 171 websites that ranged from 662 up to 789 unique webpages.

The first hypothesis and the implication that the size of a website leads to higher or lower organic search engine traffic raises a new research question. This is related with SEO technical compatibility factors, whether CHI websites follow them, and to what extent. There are prior approaches indicating that Google employs more than 250 closely guarded secret factors in its ranking algorithm [18,20]. However, quite a few studies [16,18,19,21,22] have investigated only some of them. This fact is probably

linked to the until-that-time managerial inability of the retrieval of large-scale datasets regarding technical and behavioral factors that influence organic search engine traffic percentage.

Against this backdrop, big data analytics offers new opportunities in strategic SEO planning and deployment. The data gathered through application program interface (API) integration and preprocessing, result in a plurality of technical and behavioral SEO variables that probably impact the percentage variance in the organic search engine traffic of CHI websites. Based on principal component analysis, we divided these technical and behavioral variables into four additional factors: SEO Crawling, Website Loading Speed, Website Security Condition, and User Behavior.

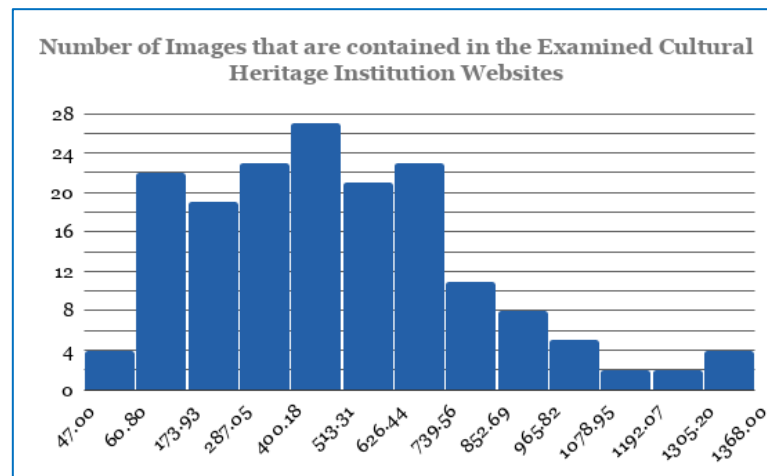


Figure 2. Histogram of the number of images that the examined CHI websites contain. The horizontal axis indicates the numerical amount of images that the examined websites contain. The vertical axis demonstrates how many websites contain from 47 up to 1368 images. It is noted that 408 up to 513 images represented the greatest number of websites, with a sample of 27 of the total 171 domains.

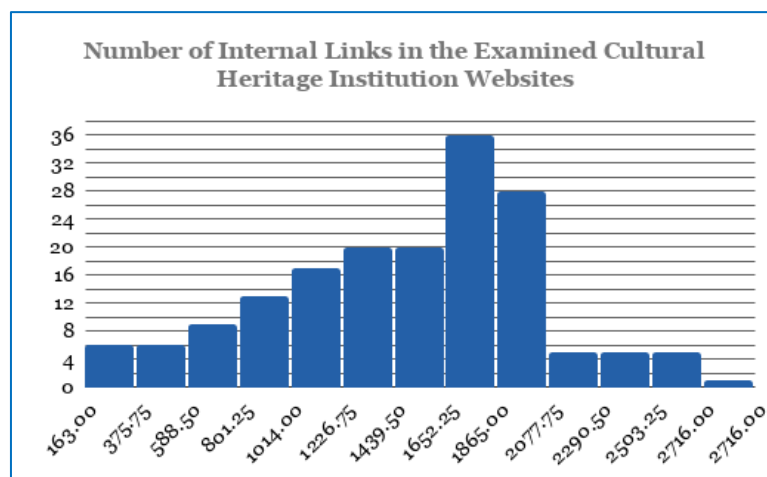


Figure 3. Histogram that explains the number of internal links included in the examined cultural heritage websites. Of the total, 36 websites ranged from 1652 up to 1865 internal links that allow users to navigate from one webpage to another with the purpose to provide cultural heritage content.

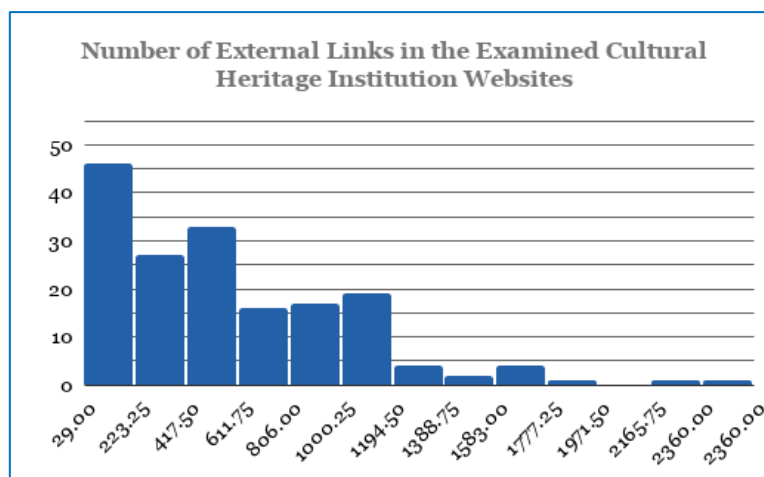


Figure 4. In contrast with the internal links, this histogram indicates that most of the cultural heritage institution websites do not contain a large number of external links inside their websites that allow users to navigate outside of them. Indeed, 45 of the total 171 websites ranged from 29 up to 223 external links.

2.2. SEO Crawling

Even from the initial approaches to identify SEO technical factors and make them comprehensible [23,24] up to the latest research implications [25–27], there are multiple de facto variables that are taken into consideration and impact the search engine rankings. These variables are related to the existence of Page Title tags [2,16,19,24,25], Meta-Descriptions [25–28], or Headings [2,23,28]. Others concentrate on aspects related to Technical Curation of Content. These are the Alt-tags on images, the prevention of duplicated page content, its re-use and repeatability, and a proper mobile scaling setup for better friendliness and adaptiveness to mobile devices [2,26,27].

Other SEO crawling variables focus on URL Diligence and their appropriate structure [25,28,29]. Link Condition, such as avoiding broken pages or setting 404 error pages for redirections [26,28], is another aspect. The Code Validity and how friendly it is to the crawlers and their indexing process is another key component. This factor investigates the availability of specific back-end files that make the crawling process more efficient, including robots.txt integration, sitemap locations, and appropriate hypertext markup language (HTML), JavaScript, and cascading style sheets (CSS) architecture [26,30–32]. These prior research efforts have demonstrated the optimization of search ranking positions.

Nevertheless, it is not yet clear whether the aforementioned factors correspond (and to what extent) to the optimization of organic search engine percent of traffic, that is, the higher the SEO technical compatibility of the websites with these variables, the higher the percentage of organic search engine traffic. Therefore, our second hypothesis is the following:

Hypothesis (H2). *The SEO Crawling factor impacts the percentage increase of organic search engine traffic in the websites of cultural heritage institutions.*

2.3. Website Loading Speed

According to Google [33], when a website improves its loading time, visitors become more engaged and their interaction with the content is increased. The number of clicks is increased and more conversions come, including sign-up forms, or purchasing products. Controversially, low loading speed could lead to reduced user–content interaction. For instance, the BBC News found that they lost up to 10% of their visitors for every additional second that their website took to load up their content [34]. On the other hand, the COOK case study as a growing business in food industry showed that, after reducing the average loading time of their website by 850 milliseconds, conversions increased

by up to 7%, pages per session increased by up to 10%, and the bounce rate decreased down to 7% [33]. Indeed, slow loading speed returns almost immediate abandonments from websites, leading users to jump to other ones with better loading speed performance.

Prior studies paid attention to some key technical issues that have significant effects on website loading speed and, therefore, on a pleasant or unpleasant navigational experience for users [2,32]. Enge and colleagues indicated the negative impact on user experience as being the longer the loading time, the shorter the visit duration and, hence, the higher the bounce rate. The key technical issues are related mostly with the compression and minification of JavaScript and CSS files and the proper architecture of link redirections inside and outside the websites [22,26,29,35–37]. Therefore, it should be examined whether the compatibility of a website with loading speed time variables is associated with an enhanced organic search engine percent of traffic. That is, the higher the percentage rates of the Website Loading Speed factor, the higher the percentage increase of organic search engine traffic. Thus, our third hypothesis unfolded as follows:

Hypothesis (H3). *The Website Loading Speed factor impacts the percentage increase of organic search engine traffic in the websites of cultural heritage institutions.*

2.4. Website Security Condition

Browsers such as Chrome or Mozilla encourage web developers to follow both fundamental and advanced security policies in their websites [38,39]. In more than a few cases, cultural heritage institutions contain online shops inside their websites. This fact demands the establishment of security technologies that foster visitor convenience to explore and buy products. In addition, visitors interact with the online cultural information and sometimes submit their personal information for further services and/or potential news and updates. From a managerial point of view, the compatibility of CHI websites with security factors ensures the reputation of the cultural organizations and prevents malware and cyberattacks that will inevitably cost economic resources for rectification and recovery.

A fundamental aspect regarding the security condition of a website is related to the existence of Hypertext Transfer Protocol Secure (HTTPS) and password and contact forms secured with reCaptcha. Other website security factors are related with the hiding of server version data, avoiding in this way the indication of what software is running on the Web server. Moreover, the use of clickjack and content sniffing protection and the settlement of Multipurpose Internet Mail Extensions (MIME) to help browsers prevent content sniffing exploits are other vital factors in terms of website security condition [39].

The practical community, such as the Open Web Application Security Project (OWASP), suggests that the aforementioned variables need to be taken into serious consideration for protecting not only websites, but also cloud-based software applications [40,41]. Nevertheless, to the best of our knowledge, there is no prior clear substantiation as to whether the website security condition impacts (and to what extent) the organic search engine percent of traffic. Apparently, there are some indications about the importance of security factors in the SEO context [22,35]. However, little is known as regards the impact of security factors on the percentage increase of organic search engine traffic. In this respect, our fourth hypothesis is the following:

Hypothesis (H4). *The Website Security Condition factor impacts the percentage increase of Organic Search Engine Traffic in the websites of cultural heritage institutions.*

2.5. User Behavior

In 2012, Agarwal and colleagues introduced a semantic schema for potential personalization of content to future users based on the behavior and interactions of prior users [42]. More specifically, they proposed that the higher the engagement with the content in terms of time of interaction and depth of exploration, the greater the content for provision to potential users. Three years later, Mavridis and

Symeonidis [20] argued that search engines try to capture user interactions inside websites in order to provide feedback and optimize ranking algorithms. In the Big Data era, a necessity arises regarding the transformation of behavioral data into semantic schemas aiming at greater personalization and optimization of the World Wide Web by the people, for the people. Indeed, analytics platforms such as Google Analytics transform behavioral patterns into semantic web analytics metrics for better visualization and clear comprehension of the total user experience inside the websites.

These Web analytics metrics were previously mentioned in prior research, including Pages per Visit [19,29,32], Visit Duration [32,35,42], Total Clicks in a specified time period, and percent of immediate abandonment of the websites through the Bounce Rate metric [2,26,34,42]. All the aforementioned prior studies significantly designated these behavioral metrics as variables that numerically evaluate the interaction and experience of users inside websites. However, clear results are missing to indicate whether User Behavior inside websites impacts any forthcoming increase of the organic search engine percent of visits. That is, the greater the interaction, the higher the percentage of organic search engine traffic. Consequently, our fifth and final hypothesis is the following:

Hypothesis (H5). *The User Behavior factor impacts the percentage increase of organic search engine traffic in the websites of cultural heritage institutions.*

In Figure 5, we present our proposed model, depicting in this way the hypothesized relationships among the factors that probably increase the organic search engine visit percent; that is, “the higher the rates of the proposed factors, the higher the percentage of search engine visits will be”. In the next section, the methodology unfolds with the purpose to answer our research questions.

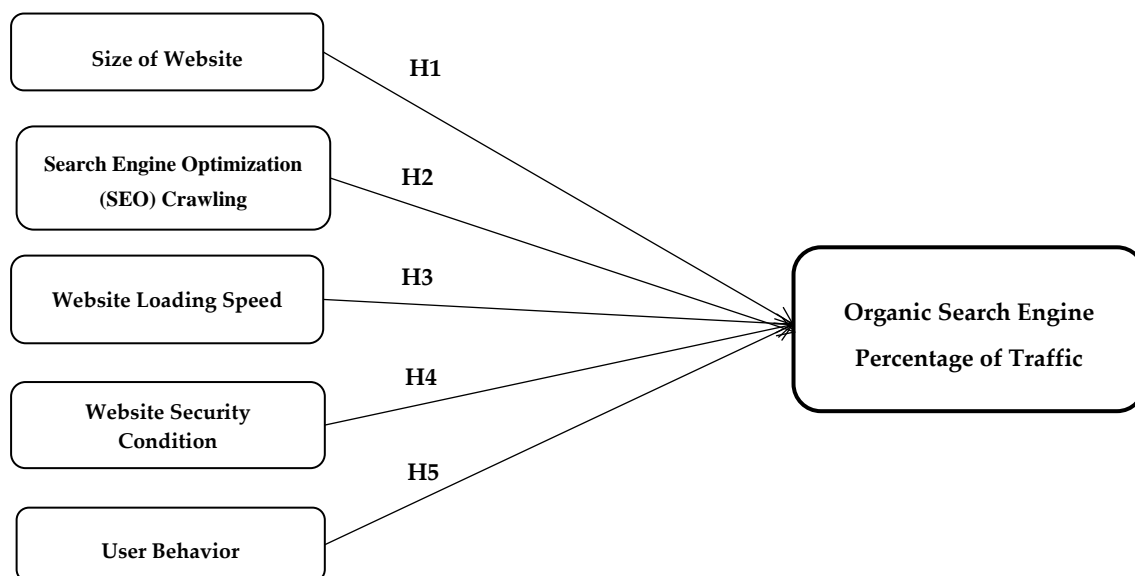


Figure 5. The proposed research model. Each of the proposed factors might have a different impact on the increase of the organic search engine percentage of traffic.

3. Methodology

The purpose of this paper is to present a novel methodology that measures and evaluates the impact of five factors, namely, the Size of Websites, SEO Crawling, Website Loading Speed, Website Security Condition, and User Behavior, on the optimization of the organic search engine percent of traffic. That is, we investigate which of these factors, and to what extent, have an impact on the number of organic visitors to CHI websites that explicitly come from search engines. In order to answer this research question, we deploy a three-stage methodology (Figure 6) composed of the following:

1. The summarization, validation, and alignment of the retrieved big data analytics with the organizations' defined key performance indicators (KPIs);
2. The development of diagnostic exploratory models that estimate the cause-and-effect relationships between the metrics;
3. The predictive model development and the process of simulation for optimization purposes.

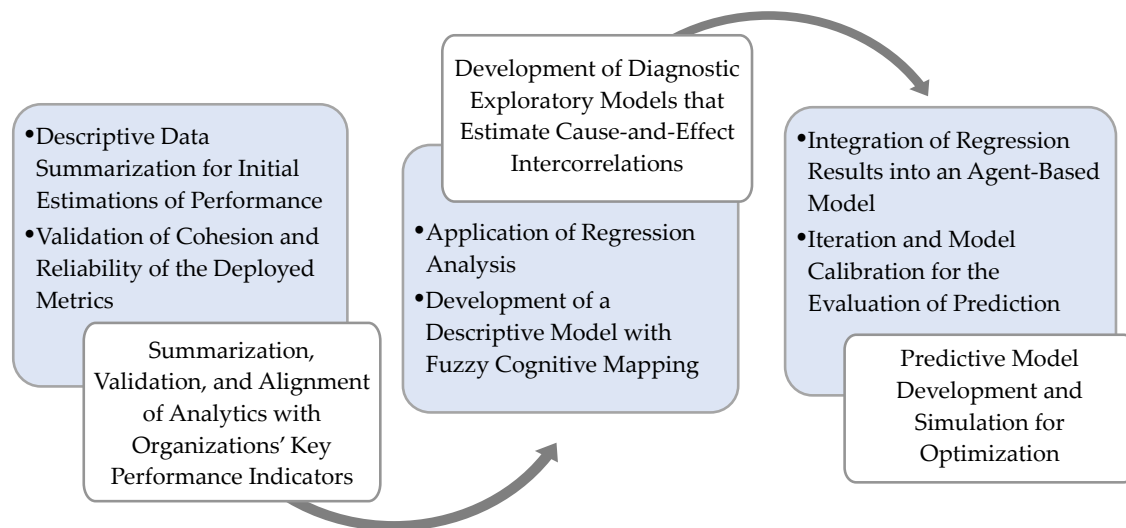


Figure 6. Representation of the proposed methodological approach. Each step supports the analysts and decision-makers of cultural institutions to retrieve, validate, organize, and utilize big data analytics in order to improve their online visibility and findability.

This methodological approach is able to tackle several issues derived from big data analytics utilization for optimization purposes. The first stage reflects efforts to cover drawbacks regarding managerial difficulties to utilize big data analytics in a more efficient way, while improving their veracity for further exploration. Moreover, the first stage corresponds to the process of aligning analytics metrics with the KPIs of cultural institutions and their Web presence. For example, in this study, the KPI is the increase of organic search engine traffic percent, which is aligned with all the Web metrics that are included in the five factors and probably affect the KPI itself.

The second stage gives the advantage of understanding through a quantified framework of evaluation the performance of a Web-based system and the possible intercorrelations between metrics. This gives the advantage of focusing on the critical ones while neglecting the less important. The third stage—inherited by the two previous stages—expands the opportunities to develop data-driven predictive models and simulations in all modeling steps. These models are composed with higher levels of validity and reliability and, hence, better precision of predictions.

3.1. Summarization, Validation, and Alignment with KPIs

In the context of big data analytics utilization, one of the most common issues is the question of whether more efficient decisions require more data or better models. This assumption returns back to the phenomena of exploring and trying to find hidden valuable gems in *big datasets* but with unfortunately *small stats*. As not all datasets are created in an equal way, we argue that neither volume of data nor better models affect the decision-making processes. Controversially, we assume that the testing process of validity, reliability, cohesion, and consistency of the extracted datasets is the first and most important thing. Besides, in the systems world, the quality of the inputs (data) always affects the outputs (models).

3.1.1. Data Sample and Preprocessing

The retrieval of big data analytics was implemented through the integration and parameterization of three different APIs. The Search Console URL API was used to retrieve data about the size of CHI

websites, incorporating for exploration the number of webpages, images, and internal and external links for each of the examined 171 domains. Google Arts & Culture Database [43] was used to capture the number of websites that were investigated. An aggregated volume of 141,851 webpages, 86,788 image path files, and 254,373 internal and 97,626 external links was retrieved from all the websites.

The Checkbot API was used to encapsulate the compatibility of technical factors, namely, SEO Crawling, Website Loading Speed, and Website Security Condition. Further parameterization was applied in the API in order to explore and discover technical patterns in the HTML, JavaScript, and CSS code that were mentioned in previous research studies but not contained in the API's initial capabilities. A scale from 0 to 100 was implemented for each one of the retrieved variables as a useful managerial approach to tackle complexity in CHI website performance measurement [2,32,44,45]. In total, 9405 unique instances were retrieved for the 55 examined variables, ranging from 0% to 100%. Additionally, we utilized the Similar Web API for the retrieval of behavioral data analytics as regards Organic Search Clicks per Month, Visit Duration, Pages per Visit, and Bounce Rate for each of the 171 CHI websites that were investigated. This kind of API was also integrated in order to gather statistics about the percentage of organic search engine traffic each website receives.

3.1.2. Validation and Reliability

Consequently, statistical analysis was implemented for validation and to ensure the reliability of the extracted variables included in each factor. This was to give a clear perception of the levels of cohesion and consistency among them. Subsequently, dimension reduction was used with principal component analysis (PCA). Preliminary test analysis was performed on the first 25 percent of all examined websites (43/171) in order to ensure that there was no kind of violation of the assumption of data normality and linearity in their nature [46,47]. A second meta-analysis was performed on the full dataset to further ensure its normality and linearity.

Furthermore, Kaiser–Meyer–Olkin (KMO) testing was performed with the purpose to ensure that the retrieved datasets were suitable for principal component analysis and potential categorization [47]. The closer the value of KMO to 1, the greater the suitability of the extracted data for PCA, and vice versa. In Table 2, the examined factors are presented with reference to the factor loading for each variable, descriptive statistics, KMO, and percent of total variance explained.

Some variables with loadings below the limit of 0.500 were dropped. These are shown with a strikethrough line. Recognizing the contribution of the statistical tests to information for decision-makers, it is noted that they are applied here with the purpose of evaluating suitability for optimizing CHI website performance and organic search engine traffic percentage.

3.2. Development of Diagnostic Exploratory Models

Linear regression was implemented for the development of diagnostic exploratory models that estimate the cause-and-effect relationship between the five factors and their impact on organic search engine percent of traffic. Through this statistical approach, decision-makers will be able to understand in a clear manner the impact of each one of the proposed factors on the total change in organic search engine traffic percentage. Moreover, it is crucial to present a practical, manageable, and comprehensible methodological framework that dynamically demonstrates the intercorrelations between the factors and the defined KPIs [44,45,48]. This will also prevent the phenomenon of “*more metrics than users*” indicated by a plethora of prior approaches inside web-based systems for visualizing performance [49]. In order to achieve this, we adopted the descriptive modeling method of Fuzzy Cognitive Mapping (FCM).

FCM is used to visualize more intuitively and effectively the relationships between factors through an understandable representation of these correlations. This can be applied through the settlement of numerical weights (of +1 or −1) for each relationship of the proposed descriptive model [50]. The ease of construction and visualization, and the ease of understanding even by non-technicians, are further advantages of adopting the FCM [51]. In addition, the deployment of the FCM as a descriptive

modeling method is characterized as a prerequisite step for the potential development of data-driven predictive models and simulations [51].

FCM provides a macro-level view of the examined system by aggregating and articulating the causal influence amongst the included entities [51,52]. More specifically, FCM demonstrates the correlations among the included factors and subfactors and how each one contributes to another [52].

This constitutes a process of feeding back to the administrators of CHI websites for them to understand in a manageable way the impact of each factor on increase or decrease in the organic search engine percent of traffic (Figure 7). Moreover, it should be noted that ranking algorithms change over time; hence, so do the factors that should be considered to increase or decrease organic search engine traffic. In this respect, FCM imparts flexibility in adding more factors and subfactors into the existed aggregated model, without new modifications that cause holistic restructuring of the descriptive situation within the model itself.

However, it is a commonly known fact that FCM constitutes an aggregated but stable macro-level approach, while missing individual and temporal explicitness at the micro-level without expressing system dynamic changes [53–55]. That is, in our case, each of the examined cultural heritage institution websites has its own performance over time relating to the technical factors and how users behave with the provided content. In order to predict and simulate possible optimization of the organic search engine percent of traffic while including temporal dynamic variances, agent-based modeling (ABM) is an appropriate method for predictive model development. Figure 7 illustrates the cause-and-effect relationships between the five proposed factors and the organic search engine traffic percent through the FCM method and its development via Mental Modeler cloud-based software.

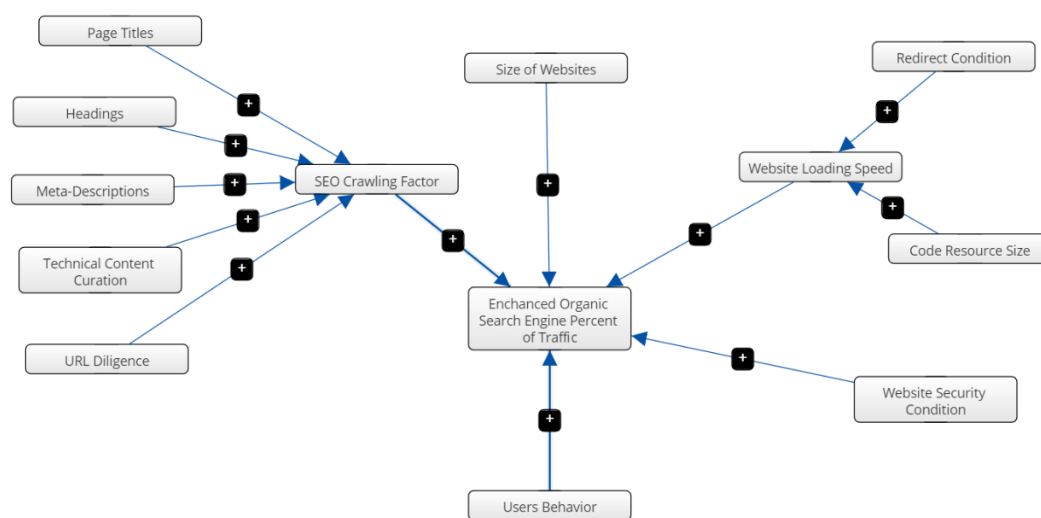


Figure 7. Fuzzy Cognitive Mapping integration as a descriptive, aggregated, and macro-level modeling method for representing the impacts among the examined factors on the percentage increase of Organic Search Engine Traffic.

3.3. Predictive Model Development and Simulation for Optimization

For the development of a predictive model and simulation run that forecasts the optimization of organic search engine traffic percent based on the impact of each factor, we proceeded with the agent-based modeling (ABM) approach. In ABM, a system is modeled as a collection of autonomous decision-making entities—the agents—while depicting relationships between them, acting via if-then rules [54,56]. The adoption of ABM as a micro-level simulation modeling strategy allows decision-makers to accomplish the following:

- Realize the full potential of the data that a cultural organization has about their online visitors and website performance. In our case, this gives a better perception of the utility of the extracted big

data analytics about the performance of websites in terms of their Size, SEO Crawling, Loading Speed, and Security compatibilities. At the same time ABM expands the opportunities to utilize behavioral datasets about user engagement with the proposed content in order to provide them with even more personalized cultural information.

- Understand the organizations—in this case, each examined CHI website—from the viewpoint of not only the services that they contain, but also activities and how users interact with the services; that is, what visitors actually do inside the cultural websites, as each visitor produces different numerical values regarding time and depth of exploration.
- Exploit the flexibility that ABM gives to organizations in terms of model construction. Highlighting its practical contribution, ABM offers the opportunity for decision-makers to construct their own model. In this way, they are able to combine entities based on the prior retrieved big datasets [56,57] and their domain knowledge and experience in the management of cultural heritage institutions. This also provides complementary flexibility to the parameterization of the model and sustainability in the entities that it contains. Indeed, as SEO topics and ranking algorithms change rapidly, ABM is more flexible in adding new factors, rather than explaining the relationships of the new factors via structured differential equations [54,58].

In the next section we present the results of the study and the integration of both the descriptive and predictive models through FCM and ABM, respectively.

4. Results

As can be seen in Table 2, most of the variables in every factor resulted in sufficient loadings. This also indicates a sufficient extent of relevance of variables in explaining the construct of each of the proposed factors. However, there were six variables that resulted in factor loadings of less than 0.500. They are depicted in the table with a strikethrough line. We decided to exclude them from the regression analysis and their involvement as independent variables in the impact that they have on the dependent variable, namely, the organic search engine visit percent. Regarding the percent of variance explained in each factor, most of them expressed considerable variability, reaching up to 55%.

Table 2. Descriptives and internal validity and consistency of the examined factors and their variables.

Factors and Items	Loading	Mean	SD	KMO	% of Variance Explained
<u>Size of Website</u>		-	-		
Pages	0.758	834.97	392.97	0.718	57.63
Images	0.884	508.31	305.14		
Internal links	0.906	1502.61	542.49		
External links	0.609	577.5	461.4		
<u>SEO Crawling</u>		72.04	38.48		
<u>Page Titles</u>	0.753	73.89	29.36	0.72	59.68
Set page titles	0.746	98.49	31.62		
Use optimal-length titles	0.845	66.57	32.26		
Use unique titles		56.63	38.67		
<u>Page Headings</u>		62	37.65		
Set H1 headings	0.912	77.64	32.26	0.679	66.85
Use one H1 heading per page	0.755	60.64	38.67		
Use optimal-length H1 headings	0.883	73.84	31.97		
Use unique H1 headings	0.702	36.88	33.45		
<u>Page Meta-Descriptions</u>		32.09	36.06		
Set page meta-descriptions	0.903	48.32	40.6	0.707	63.91
Use optimal-length meta-descriptions	0.843	26.08	31.62		
Use unique meta-descriptions	0.714				
		23.32	29.97		

Table 2. Cont.

Factors and Items	Loading	Mean	SD	KMO	% of Variance Explained
<i>Page Content Technical Curation</i>		81.26	31.92		
Avoid thin-content pages					
Set image ALT text	0.778	78.57	29.78	0.721	68.87
Set mobile scaling	0.687	62.9	42.21		
Avoid plugins	0.839	87.61	30.47		
Avoid duplicate page content	0.727	98.59	8.25		
	0.881	78.84	27.12		
<i>URL Diligence</i>		79.5	32.04		
Use short URLs	0.795	83.52	21.47		
Avoid URL extensions	0.753	80.8	35.64		
Avoid URL parameters	0.665	76.89	27.71	0.679	59.14
Avoid symbols in URLs	0.669	95.41	16.21		
Set canonical URLs	0.721	28.25	43.01		
Use lowercase URLs	0.487	87.25	27.26		
Avoid underscores in URLs	0.637	90.43	23.68		
Avoid deeply nested URLs	0.288	94.43	16.09		
<i>Link Condition</i>		80.06	35.05		
Use 404 code for broken URLs	0.847	85.76	33.39		
Avoid broken internal links	0.663	73.22	37.92	0.754	60.02
Avoid broken external links	0.555	79.24	35.41		
Avoid broken page resources	0.711	83.7	32.77		
<i>Code Validity and Crawling</i>		69.8	41.86		
Valid HTML	0.762	30.67	39.66		
Valid CSS	0.935	98.46	4.2		
Valid JavaScript	0.712	99.19	7.72		
Avoid excessive inline JavaScript	0.579	83.8	32.26	0.692	64.8
Avoid render-blocking JavaScript	0.701	18.28	36.88		
Avoid excessive inline CSS	0.547	96.36	18.08		
Avoid CSS @import	0.74	92.39	19.72		
Use of robots.txt file	0.759	83.17	35.82		
Use of Sitemap locations	0.874	25.23	43.24		
<i>Website Loading Speed</i>		79.11	34.61		
<i>Code Resource Size</i>		82.76	30.18		
Use compression	0.816	78.18	39.08	0.728	56.11
Avoid recompressing data	0.372	96.83	15.89		
Use minification	0.764	57.88	24.85		
Avoid inline source maps	0.542	98.23	13.12		
<i>Redirect Conditions</i>		76.19	37.54		
Avoid internal link redirects	0.652	42.49	45.06		
Avoid temporary redirects	0.557	70.28	36.05	0.686	44.18
Avoid redirect chains	0.803	91.01	14.69		
Avoid meta redirects	0.614	92.08	26.65		
Avoid resource redirects	0.44	84.85	33.01		
<i>Website Security</i>		53.57	43.52		
Use HTTPS	0.891	83.08	37.17		
Avoid mixed content	0.385	94.26	17.58		
Use secure password forms	0.584	92.89	24.81		
Use HSTS	0.796	11.09	31.47	0.781	64.96
Use HSTS preload	0.621	2.31	15		
Use XSS protection	0.663	15.83	36.3		
Use content sniffing protection	0.86	32.66	41.92		
Set MIME types					
Use clickjack protection	0.823	99.26	7.64		
Hide server version data	0.877	49.33	49		
	0.324	55.74	45.8		

Table 2. Cont.

Factors and Items	Loading	Mean	SD	KMO	% of Variance Explained
<u>User Behavior</u>		-	-		
Organic Clicks per Month	0.968	266,851.00	71,147.00	0.822	69.53
Visit Duration	0.913	2.58	1.16		
Pages per Visit	0.884	3.85	1.94		
Bounce Rate	0.805	55.54	9.88		
<i>Organic Search Engine Percent of Traffic</i>	<i>N: 171</i>	<i>Min: 15.26</i>	<i>Max: 89.80</i>	<i>Mean: 62.76</i>	<i>SD: 13.53</i>

The underlines depict the factor and its items, the italics depict the sub-factors, and the strikethrough lines depict each dropped variable.

Linear regression analysis returned significant indications (Table 3). Prior preprocessing and analysis was conducted in order to exclude outliers that would possibly influence the outcomes of the prediction. We also note that no changes were observed when we used a hierarchical technique. From the results of the test, all the produced models have clear statistical significance with p values less than 0.05. Nevertheless, there is a marginal exception in the factor of Website Loading Speed with a p value of 0.061, greater than 0.05. In terms of R^2 values, the results depict alignment with prior studies stating that ranking algorithms involve a massive number of secret variables in the SEO context [20,48]. In fact, Size of Websites explained up to 30.06% of the variability of the response data around its mean; SEO Crawling, up to 17.7%; Website Loading Speed, up to 10.01%; Website Security Condition, up to 17.9%; and User Behavior, up to 29.5%.

Table 3. Regression analysis output.

Variable	Coefficient	R ²	F	p Value
<i>Constant (Organic Search Engine Traffic Percent)</i>	62.071	0.306	2.827	0.012 *
Size of Website	1.060			
<i>Constant</i>	77.061	0.177	5.458	0.021 *
SEO Crawling	2.141			
<i>Constant</i>	73.301	0.101	1.855	0.061
Website Loading Speed	1.115			
<i>Constant</i>	71.702	0.189	5.589	0.019 *
Website Security Condition	1.361			
<i>Constant</i>	71.309	0.295	7.497	0.000 **
User Behavior	3.141			

$N = 171$. * and ** indicate statistical significance at the 95% and 99% levels, respectively.

Indeed, the more numerous the variables, the higher the model fit and, consequently, the R^2 values [59]. In other words, we defined some variables in each factor; however, all of them require further research to investigate even more variables that play a crucial role in the increase of the organic search engine traffic percent.

In terms of the first hypothesis, a significant regression equation was found with $p = 0.012$ and R^2 of 0.306. This means that the mean value of the organic search traffic percent is increased by 1.06% for every percentage point increase in each variable of the Size of Websites factor. For the second hypothesis, a significant regression equation was observed with $p = 0.021$ and R^2 of 0.177. This means that the mean value of the organic search traffic percent is increased by 2.14% for every percentage point increase in the percent of the SEO Crawling factor.

For the third hypothesis, a slightly nonsignificant regression equation was found with a marginal value of $p = 0.061$ and R^2 of 0.101. Even by bootstrapping of 1000 additional re-samples, there was no significant change in the p value, R^2 , or coefficients. Therefore, the mean value of the organic search traffic percent is increased by 1.11% for every percentage point increase in the Website Loading Speed

factor. For the fourth hypothesis, Website Security Condition provided a significant regression equation with $p = 0.019$ and R^2 of 0.189. As a result, the mean value of the organic search traffic percentage is increased by 1.36% for every unit increase in the Website Security Condition factor.

For the last hypothesis, User Behavior in the examined cultural websites seems to have the highest impact on the increase in organic search engine traffic percent. More specifically, a significant regression equation was observed with $p = 0.000$ and R^2 of 0.295. Thus, the mean value of the organic search traffic percent is increased by 3.14% for every percentage point increase in each variable of the User Behavior factor. This constitutes an important research result. Prior research implied that search engines encapsulate user behavior in order to rank websites in their search engine result pages [20,48]. And indeed, the higher the ranking position of a website, the higher the organic search engine traffic percentage that they receive [18,19,60].

Agent-Based Model Development

The extracted outcomes of the regression statistics resulted in significant implications that could be incorporated into a predictive data-driven agent-based model. The purpose of ABM is to compute and represent each case individually at a micro-level view while taking into consideration temporal changes [53–55]. This is not possible through the aggregated macro-level approach of Fuzzy Cognitive Mapping.

For instance, in ABM, decision-makers are able to estimate the impact of each change in SEO performance and user behavior individually for each website. This advantage provides precise results regarding the impact that each SEO factor has for every website and also about the percentage variance of organic search traffic that the website receives. As all cultural websites differ in their content and how compatible they are with the SEO factors, it is possible to generate different data analytics, both technical and behavioral. This means that managerial staff need more or less time to rectify SEO issues and, therefore, to improve user behavior and enhance organic search traffic. In this case, ABM as a predictive modeling and simulation method gives risk-free flexibility to decision-makers. They are able to themselves define the time needed to keep up with the SEO compatibility framework and, thereafter, to improve their organic search traffic. This approach combines both managers' domain knowledge, as each cultural organization differs in its operations, and the practical insights of the data analytics results.

The first main goal of the proposed computational model is verification that the proposed methods and results can be used to optimize the organic search engine traffic percentage. The second goal is decrease of the bounce rate level as a negative indicator of the overall user behavior in the examined cultural heritage websites. In the next figure (Figure 8), we present the developed ABM, its entities, and its conditions. For model development, AnyLogic ver. 8.5.2 software was used in JAVA source code in order to compute agent behavior. We defined a specific time range of 90 days in order to predict and simulate the percentage increase of organic search engine traffic. No additional days were included as there was no other crucial percentage variance in the organic search engine and bounce rates.

The model starts in its initial stage with the possibility of entrance and visit inside the cultural heritage institution websites, an initial point that depicts fundamental aspects of the agent-based development process [54–56]. This is indicated in the first statechart, entitled "Potential Search Engine Users". The transition of users (as agents) among statecharts is computed based on the prior descriptive statistics of the study, such as min, max, mean, and mode, and the outcomes of the regression analysis. The impact level that users receive from the Size of Websites, the Website Loading Speed, and the Website Security is defined by the conditions of the Size of Websites Impact, Website Loading Speed Impact, and Website Security Impact. These three major factors and their defined conditions impact drastically on user behavior inside cultural heritage websites. However, as the regression results indicated, the Website Loading Speed factor does not impact the organic traffic percentage, so there is no kind of transition between the two statecharts. At the same time, the Size of Websites, Website

Security, and SEO Crawling factors are depicted as statecharts that impact both User Behavior and the final goal, which is positive influence of the Organic Search Engine Traffic Percentage.

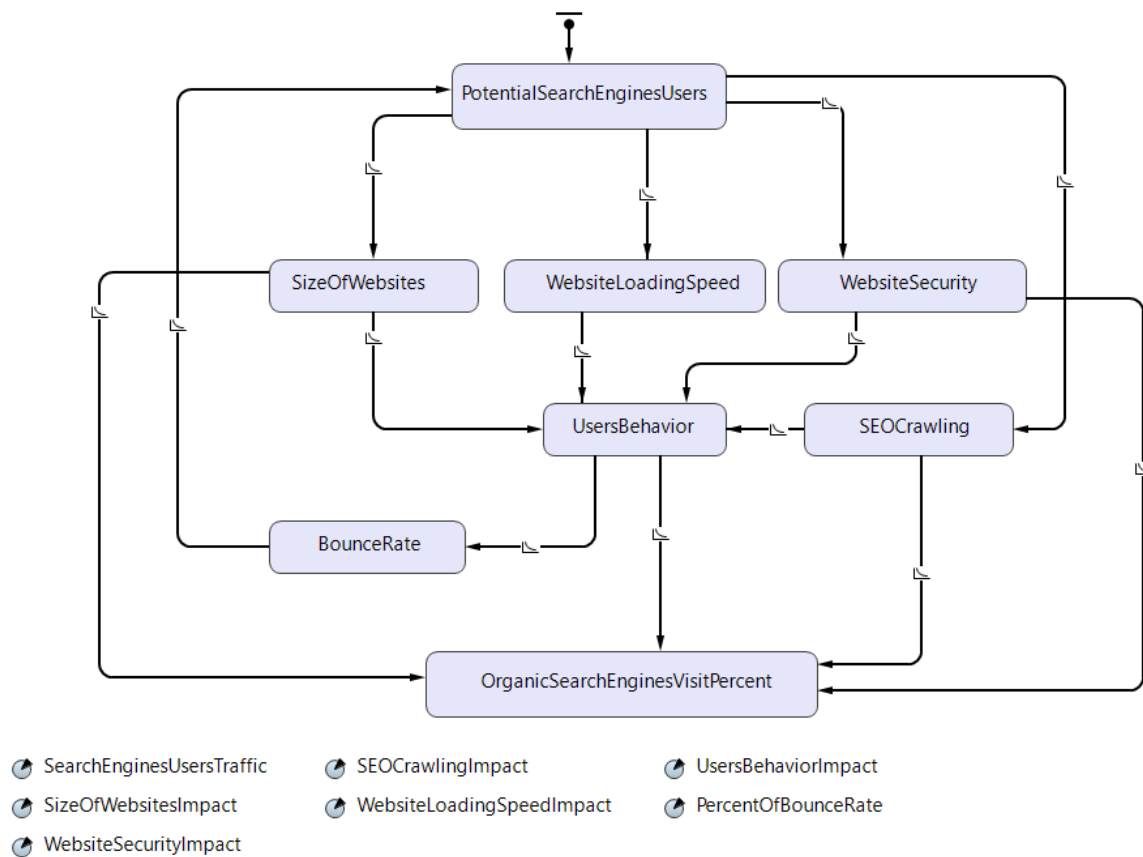


Figure 8. A predictive agent-based model for optimization of the organic search engine visit percentage and decrease of the bounce rate.

This kind of impact is computed through Poisson distributions, while setting up the results of the regression as lambda values (λ). The Poisson distribution was selected because our sample

- highlights predictive optimization in the specific time range of 90 for which the model runs;
- is generated by a known constant mean rate of values, most of them with the simplicity of a percentage; and,
- each one of the examined websites and their performance is completely independent and is not affected by the previous performances of the other websites.

Moreover, we define the consequence of low interaction rate and dissatisfied user behavior resulting in immediate abandonments via the bounce rate metric, which is illustrated through the Bounce Rate statechart and computed through the Percent of Bounce Rate condition. That is, users enter the websites, but they find insufficient content (Size of Websites), low speed of navigation (Website Loading Speed), and insufficient security in their exploration, so they leave the websites almost immediately upon their visit. In Figure 9 we present the outcomes of the predictive agent-based model.

The graph in Figure 9 represents the potential scenario of improving each of the examined factors based on the outcomes of the regression and their impact on the organic search engine traffic percentage and bounce rate level. Indeed, the model, after the initial days of the run, shows an improvement in organic search engine traffic. At the same time, a decrease and steadiness is observed in the bounce rate level without any significant sign of increase. Furthermore, it is noted that the Organic Search Engine Visit Percent does not show any further optimization after Day 50, following a straight line

without any kind of fluctuation or change. This happens for two reasons. First the examined cultural websites are able to receive the rest of their visitors from paid advertising, social networks, direct traffic, email marketing campaigns, and other external websites. Secondly, it is possible for the managerial staff to cover and rectify all the SEO technical compatibility factors that are able to optimize the organic search traffic up to ~75%, as depicted in Figure 9. Therefore, new data analytics and regression results are needed in order to provide feedback, update the predictive model, and determine the potential extent of the percentage increase in organic search engine traffic.

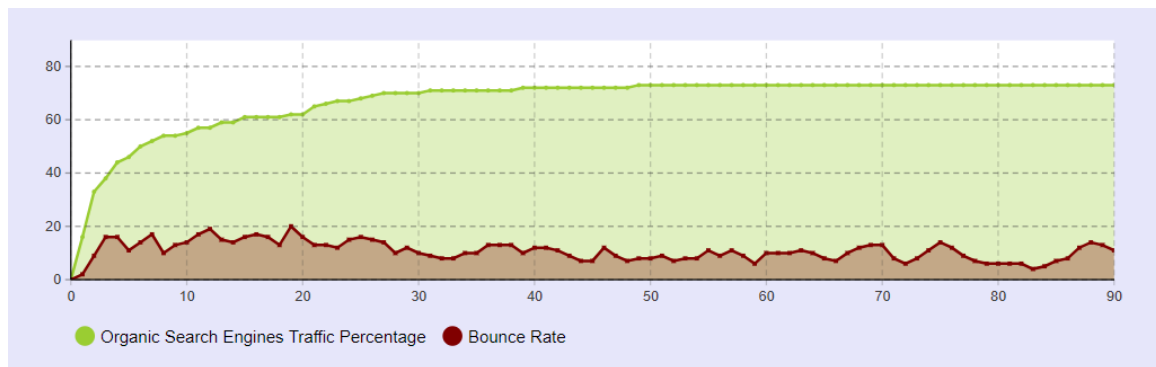


Figure 9. Optimization of the organic search engine traffic percentage in a time range of 90 sequential days. The horizontal axis demonstrates the specific time range from 0 up to 90 days of the simulation run. The vertical axis depicts the percentage of organic search engine traffic.

5. Discussion and Future Implications

The optimization of visibility in cultural heritage websites improves the knowledge that stakeholders receive from them. More specifically, in an open and democratized way, CHI websites increase people's interest in the past and allow them to recognize how their surrounding societal context has changed over time. In this respect, the SEO strategy must be set within the prism of the overall mission of cultural heritage institutions, rather than assigning these functionalities unilaterally to technical staff. This will enhance the overall importance of SEO strategies from the support personnel to the upper management levels.

In this paper, we proposed a novel methodology that quantifies in a manageable way the impact of several factors on the organic search engine traffic percentage in CHI websites with the purpose of increasing visibility and findability. One step further, this methodology offers to the administrators of CHI websites an opportunity to convert data analytics about SEO performance into useful insights and actions for potential optimization of their visibility on the Web. Otherwise, a big data analytics framework without evaluation, analysis, interpretation, and suggestion for further improvement is completely useless [61]. Based on that, we believe that this research establishes a new SEO context of communication, as more and more big data analytics can be retrieved and interpreted, while focusing on critical factors and omitting less relevant ones in organic traffic optimization. In this respect, this methodology provides new opportunities both for managers in cultural institutions and for research on this topic.

5.1. Managerial Implications for Cultural Heritage Institutions

5.1.1. Implications for Optimized Website Performance and Visibility

The proposed model shows validity, reliability, and cohesion as regards the variables and factors that it contains for evaluation and, hence, optimization of organic search engine traffic. This works as a solid stepping stone for managers to adopt this methodology, evaluating the importance of each factor and focusing precisely on each one for further improvement. However, although the results of the

obtained behavioral data analytics demonstrated that CHI websites receive on average up to 62.76% of their total traffic solely from search engines (Table 2), the bounce rate level was observed to reach up to 55%. This means that more than one in two visitors immediately abandon the websites after their visit.

Based on that, we suggest to marketing managers and administrators of cultural websites to focus first on usability improvement with the purpose to improve user interaction and behavior. The factor of SEO Crawling includes variables that might have crucial impacts for enhanced usability and user experience in CHI websites. The avoidance of thin or duplicated content, removal of broken links that confuse users, and proper mobile friendliness are some of these user-centered variables. Moreover, this factor includes variables that have a systemic impact on increasing search engine friendliness to crawlers. The appropriate curation of headings, titles, meta-descriptions, robots.txt, and sitemap files are some of the points that managers should focus on with the purpose to develop favorable conditions for the indexing process by search engines.

In this research, regression results show that user behavior has the highest impact among the factors that affect the percentage increase of organic search engine traffic (Table 3). Indeed, User Behavior can increase by up to 3.14% the total percentage of organic search traffic in the examined websites. Nevertheless, if the administrators do not pay greater attention to aligning their efforts firstly with the optimization of website usability, then the behavior of users and their experience will negatively affect the percentage of search engine traffic. Therefore, the managerial staff of cultural institutions must not focus only on SEO strategies that aim for ranking optimization; it is more important to improve usability for better engagement between users and content [60,62]. This will positively increase behavior and, thereafter, provide higher organic traffic percentages.

5.1.2. Utility of the Methodology

The proposed methodology not only supports managerial staff to seek an aggregated evaluation of a CHI website, it is also a flexible approach to focus on the individual performance of specific collections contained in unique webpages which unfortunately suffer from low visibility and findability on the Web. For instance, administrators could evaluate the SEO performance of specific webpages of cultural content while manipulating in a more efficient way the process of optimization in specific *parts* rather than the *whole*. This approach covers a rigorous and challenging task for cultural heritage institutions, as they have to deal with the large size of their content. Indeed, the larger the size of a web-based system, the more complex its manipulations and rectifications [12,19,45,54].

In addition, the outcomes of this methodology provide practical and educational implications for cultural institutions to avoid big data frameworks that rely more on data storage and not enough on their analysis. As the reliability of the gathered data constitutes a core value for the quality of a well-informed decision-making process [61], website analysts of cultural heritage websites should focus more on big data metrics systems that fit the following:

- (a) Are aligned with their key performance indicators, such as improvement of organic search engine traffic or user engagement with the content [44,48]; and
- (b) Gather reliable and veracious data that can be deployed in predictive model development and calibration [57,63].

Therefore, the proposed methodology offers the flexibility to tackle other problematic issues in the online presence of cultural heritage institutions, such as the proper utilization of analytics for social media optimization or cost-effective online paid advertising campaigns. That is, we proceed to the careful delimitation of KPIs; gather, validate, and examine the data analytics that align with the KPIs; and then develop data-driven predictive models for optimization.

5.1.3. Optimized Financing Resource Management

In the EU context, the Eurostat report of 2018 [64] depicted a low percentage of expenditures in cultural services ranging from 0.4% up to 0.8% of the GDP. Bearing in mind the reduced financial

flexibility of cultural institutions and their limited available resources for the management of the cultural material that they contain, the process of search engine optimization could be a cost-effective marketing strategy.

In contrast with other digital marketing strategies that increase website visitor numbers but provide poor content curation and usability, SEO constitutes a sustainable digital marketing strategy that focuses on one of the most fundamental aspects of digital marketing: the effectiveness of the landing page. If users *land* in webpages that express usability and proper curation of content, then their experience will be better, making any kind of marketing communication strategy more effective. This constitutes a promising approach to reduce paid advertising strategies that do not eventually return the investment, due to the minimal interaction of users after visiting websites.

5.2. Research Implications

The dimension reduction results through principal component analysis indicated that most of these variables are suitable for providing reliable evaluations of website performance and the impact that they have on user behavior and organic search engine traffic. Notwithstanding, even if we include more than 50 variables that impact the organic search engine traffic optimization, the regression R-square values indicate the depth of search engine ranking algorithms and the multidimensionality of the variables they involve. Following the above findings, we have started further research efforts in order to explore and include more variables or factors that probably influence rankings and, hence, percentage variance in organic search engine visits. Based on this assumption, as big data mining and analytics techniques are getting more and more sophisticated and impact organizations' decision-making process in terms of marketing and promotion strategies [62], the research topic of SEO reliance in big data analytics will be discussed in a detailed manner in the future.

Regarding the predictive agent-based model development, from the initial research approaches as a computational method to describe the complexity of a system and its entities, to recent characterizations, it is referred to more as an art than as a science [56,58,65]. However, as big data analytics expand the opportunities for integrating more and more data into simulation models, the *art* is sidelined. Therefore, new research approaches are developed to overcome a lack of data and combine prior domain knowledge and analytics for logical well-informed and data-driven predictive models.

In this paper, we developed the ABM as a supportive tool that provides feedback to the managers of cultural heritage institutions regarding the impact of several factors on user behavior and organic search engine traffic percent. An abstraction level that describes the impact of each factor was developed. Nevertheless, predictive models are mostly stable in abstraction levels but unstable with larger perturbations when more conditions and entities are included. Therefore, further research is needed to evaluate predictive model efficiency through ABM when expanding the level of abstraction or integrating system dynamics approaches [66].

Furthermore, the results of the study emphasize the necessity to redefine the SEO topic. Apparently, the higher the compatibility of the SEO factors, the higher the rankings and search engine visit percentage. However, the main aim of search engines is confirmation that they provide the most qualitative content, in the highest volume, in the fastest time to their users, according to their search terms [67]. In this respect, Web developers and content creators should have practical quantified indicators in order to evaluate and optimize their website performance and content. Thus, we redefine SEO as it is not solely the process of rectification for higher search rankings. It is rather much more a *user-centric* strategy that improves the findability and visibility of information in search results and aims for integrated user experience inside websites.

Author Contributions: All the authors have contributed equally in this research effort in terms of the conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing the original draft and writing review-editing. D.K.-M., G.A.G. and D.P.S. contributed to the supervision and project administration of this study. All authors have read and agreed to the published version of the manuscript

Funding: This research received no external funding.

Acknowledgments: Dedicated to those who fought against the covid-19, but did not make it. We heartily thank the guest editors of *Big Data Analytics in Cultural Heritage* for giving us this opportunity. We are also grateful to the Reviewers for their vital remarks. To conclude, we acknowledge the valuable technical contribution of Nikolaos Lazaridis for the provision of cutting-edge hardware helping us to work faster for this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Di Franco, P.D.G.; Matthews, J.L.; Matlock, T. Framing the past: How virtual experience affects bodily description of artefacts. *J. Cult. Herit.* **2016**, *17*, 179–187. [CrossRef]
2. Krstić, N.; Masliković, D. Pain points of cultural institutions in search visibility: The case of Serbia. *Libr. Hi Tech News* **2018**, *37*, 496–512. [CrossRef]
3. Sarraf, S. A survey of museums on the web: Who uses museum websites? *Curator Mus. J.* **1999**, *42*, 231–243. [CrossRef]
4. Lykourantzou, I.; Antoniou, A. Digital innovation for cultural heritage: Lessons from the european year of cultural heritage. *SCIRES Sci. Res. Inf. Technol.* **2019**, *9*, 91–98. [CrossRef]
5. Wilson, D.M. What do we need money for? In *Money, Money, Money, and Museums*; Scottish Museums Council: Edinburgh, UK, 1991; p. 11.
6. Hume, M.; Mills, M. Building the sustainable iMuseum: Is the virtual museum leaving our museums virtually empty? *Int. J. Nonprofit Volunt. Sect. Mark.* **2011**, *16*, 275–289. [CrossRef]
7. Chiang, H.H.; Tsaih, R.H.; Han, T.S. Measurement development of service quality for museum websites displaying artifacts. In *Managing Innovation and Cultural Management in the Digital Era*; Routledge: Abingdon, UK, 2012; pp. 100–127.
8. Fantoni, S.F.; Stein, R.; Bowman, G. Exploring the relationship between visitor motivation and engagement in online museum audiences. *Museums and the Web*; San Diego, CA, USA, 11–14 April 2012. Available online: https://www.museumsandtheweb.com/mw2012/papers/exploring_the_relationship_between_visitor_mot (accessed on 2 April 2020).
9. Voorbij, H. The use of web statistics in cultural heritage institutions. *Perform. Meas. Metr.* **2010**, *11*, 266–279. [CrossRef]
10. Marty, P.F. Museum websites and museum visitors: Before and after the museum visit. *Mus. Manag. Curatorship* **2007**, *22*, 337–360. [CrossRef]
11. Ardissono, L.; Kuflik, T.; Petrelli, D. Personalization in cultural heritage: The road travelled and the one ahead. *User Model. User Adapt. Interact.* **2012**, *22*, 73–99. [CrossRef]
12. Davoli, P.; Mazzoni, F.; Corradini, E. Quality assesment of cultural web sites with fuzzy operators. *J. Comput. Inf. Syst.* **2005**, *46*, 44–57. [CrossRef]
13. Schmitz, P.L.; Black, M.T. The Delphi toolkit: Enabling semantic search for museum collections. In Proceedings of the International Conference for Culture and Heritage Online, Torondo, CA, USA, 31 March 2008; Available online: <http://www.archimuse.com/mw2008/papers/schmitz/schmitz.html> (accessed on 2 April 2020).
14. Salah, A.A.; Manovich, L.; Salah, A.A.; Chow, J. Combining cultural analytics and networks analysis: Studying a social network site with user-generated content. *J. Broadcast. Electr. Media* **2013**, *57*, 409–426. [CrossRef]
15. Kaushik, A. *Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
16. Gandour, A.; Regolini, A. Web site search engine optimization: A case study of FragforNet. *Libr. Hi Tech News* **2011**, *28*, 6–13. [CrossRef]
17. Ryan, D.; Jones, C. *Digital Marketing: Marketing Strategies for Engaging the Digital Generation*; Kogan Page Ltd.: London, UK; Philadelphi, PA, USA, 2009.
18. Luh, C.J.; Yang, S.A.; Huang, T.L.D. Estimating Google’s search engine ranking function from a search engine optimization perspective. *Online Inf. Rev.* **2016**, *40*, 239–255. [CrossRef]
19. Moreno, L.; Martinez, P. Overlapping factors in search engine optimization and web accessibility. *Online Inf. Rev.* **2013**, *37*, 564–580. [CrossRef]
20. Mavridis, T.; Symeonidis, A.L. Identifying valid search engine ranking factors in a Web 2.0 and Web 3.0 context for building efficient SEO mechanisms. *Eng. Appl. Artif. Intell.* **2015**, *41*, 75–91. [CrossRef]

21. Weideman, M.; Visser, E.B. Fusing website usability and search engine optimisation. *S. Afr. J. Inf. Manag.* **2014**, *16*, 1–9. [CrossRef]
22. Omprakash, K.S. Concept of search engine optimization in web search engine. *Int. J. Adv. Eng. Res. Stud.* **2011**, *1*, 235–237.
23. Chambers, R. Application of best practice towards improving Web site visibility to search engines: A pilot study. *S. Afr. J. Inf. Manag.* **2005**, *7*, 1.
24. Zhang, J.; Dimitroff, A. The impact of metadata implementation on webpage visibility in search engine results (Part II). *Inf. Process. Manag.* **2005**, *41*, 691–715. [CrossRef]
25. Mittal, A.; Sridaran, R. Evaluation of Websites' Performance and Search Engine Optimization: A Case Study of 10 Indian University Websites. In Proceedings of the 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 13–15 March 2019; pp. 1227–1231.
26. Salminen, J.; Corporan, J.; Marttila, R.; Salenius, T.; Jansen, B.J. Using Machine Learning to Predict Ranking of Webpages in the Gift Industry: Factors for Search-Engine Optimization. In Proceedings of the 9th International Conference on Information Systems and Technologies, Cairo, Egypt, 24–26 March 2019; Volume 6, pp. 1–8. [CrossRef]
27. Gek, D.; Kukartsev, V.; Tynchenko, V.; Bondarev, A.; Pokushko, M.; Dalisova, N. The problem of SEO promotion for the organization's web representation. *SHS Web Conf.* **2019**, *69*, 00122. [CrossRef]
28. Khraim, H.S. The impact of search engine optimization on online advertisement: The case of companies using E-Marketing in Jordan. *Am. J. Bus. Manag.* **2015**, *4*, 76–84. [CrossRef]
29. Baye, M.R.; De los Santos, B.; Wildenbeest, M.R. Search engine optimization: What drives organic traffic to retail sites? *J. Econ. Manag. Strateg.* **2016**, *25*, 6–31. [CrossRef]
30. Shih, B.Y.; Chen, C.Y.; Chen, Z.S. Retracted: An empirical study of an internet marketing strategy for search engine optimization. *Hum. Factors Ergon. Manuf. Serv. Ind.* **2013**, *23*, 528–540. [CrossRef]
31. Dawson, A.; Hamilton, V. Optimising metadata to make high-value content more accessible to Google users. *J. Doc.* **2006**, *62*, 307–327. [CrossRef]
32. Al-Ananbeh, A.A.; Ata, B.A.; Al-Kabi, M.; Alsmadi, I. Website usability evaluation and search engine optimization for eighty Arab university websites. *Basic Sci. Eng.* **2012**, *21*, 107–122.
33. Why Performance Matters. Web Fundamentals. Available online: <https://developers.google.com/web/fundamentals/performance/why-performance-matters> (accessed on 4 March 2020).
34. How the BBC Builds Websites That Scale. Available online: <https://developers.google.com/web/fundamentals/performance/why-performance-matters> (accessed on 4 March 2020).
35. Enge, E.; Spencer, S.; Stricchiola, J.; Fishkin, R. *The Art of SEO*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2012.
36. Malaga, R.A. Worst practices in search engine optimization. *Commun. ACM* **2008**, *51*, 147–150. [CrossRef]
37. Egri, G.; Bayrak, C. The role of search engine optimization on keeping the user on the site. *Procedia Comput. Sci.* **2014**, *36*, 335–342. [CrossRef]
38. Communicating the Dangers of Non-Secure HTTP. Available online: <https://blog.mozilla.org/security/2017/01/20/communicating-the-dangers-of-non-secure-http/> (accessed on 4 March 2020).
39. Why HTTPS Matters. Web Fundamentals. Available online: <https://developers.google.com/web/fundamentals/security/encrypt-in-transit/why-https> (accessed on 4 March 2020).
40. Modern Web Security Patterns OWASP. Available online: <https://owasp.org/www-chapter-sacramento/assets/slides/20200117-modern-web-security-patterns.pdf> (accessed on 4 March 2020).
41. Nigam, K.; Saxena, S.; Gupta, N. An analysis on improvement of website ranking using Joomla. *IITM J. Manag. IT* **2015**, *6*, 69–72.
42. Agarwal, S.; Nishar, D.; Rubin, A.E. Providing Digital Content Based on Expected User Behavior. U.S. Patent US 8,271,413 B2, 18 September 2012.
43. Google Arts & Culture. Available online: <https://artsandculture.google.com/> (accessed on 23 March 2020).
44. Saura, J.R.; Palos-Sánchez, P.; Cerdá Suárez, L.M. Understanding the digital marketing environment with KPIs and web analytics. *Future Internet* **2017**, *9*, 76. [CrossRef]
45. García, M.D.M.R.; García-Nieto, J.; Aldana-Montes, J.F. An ontology-based data integration approach for web analytics in e-commerce. *Expert Syst. Appl.* **2016**, *63*, 20–34. [CrossRef]
46. McFee, B.; Lanckriet, G.R. Metric learning to rank. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 775–782.

47. Dziuban, C.D.; Shirkey, E.C. When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychol. Bull.* **1974**, *81*, 358–361. [\[CrossRef\]](#)
48. Nakatani, K.; Chuang, T.T. A web analytics tool selection method: An analytical hierarchy process approach. *Internet Res.* **2011**, *21*, 171–186. [\[CrossRef\]](#)
49. Welling, R.; White, L. Web site performance measurement: Promise and reality. *Manag. Serv. Qual. Int. J.* **2006**, *16*, 654–670. [\[CrossRef\]](#)
50. Papageorgiou, E.I.; Salmeron, J.L. Learning fuzzy grey cognitive maps using nonlinear hebbian-based approach. *Int. J. Approx. Reason.* **2012**, *53*, 54–65. [\[CrossRef\]](#)
51. Glykas, M. *Fuzzy Cognitive Maps: Advances in Theory, Methodologies, Tools and Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010; Volume 247.
52. Van Vliet, M.; Kok, K.; Veldkamp, T. Linking stakeholders and modellers in scenario studies: The use of Fuzzy Cognitive Maps as a communication and learning tool. *Futures* **2010**, *42*, 1–14. [\[CrossRef\]](#)
53. Giabbanelli, P.J.; Gray, S.A.; Aminpour, P. Combining fuzzy cognitive maps with agent-based modeling: Frameworks and pitfalls of a powerful hybrid modeling approach to understand human-environment interactions. *Environ. Model. Softw.* **2017**, *95*, 320–325. [\[CrossRef\]](#)
54. An, L. Modeling human decisions in coupled human and natural systems: Review of agent-based models. *Ecol. Model.* **2012**, *229*, 25–36. [\[CrossRef\]](#)
55. Mehryar, S.; Sliuzas, R.; Schwarz, N.; Sharifi, A.; van Maarseveen, M. From individual Fuzzy Cognitive Maps to Agent Based Models: Modeling multi-factorial and multi-stakeholder decision-making for water scarcity. *J. Environ. Manag.* **2019**, *250*, 109482. [\[CrossRef\]](#)
56. Bonabeau, E. Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci. USA* **2002**, *99* (Suppl. 3), 7280–7287. [\[CrossRef\]](#)
57. Kavak, H.; Padilla, J.J.; Lynch, C.J.; Diallo, S.Y. Big data, agents, and machine learning: Towards a data-driven agent-based modeling approach. In *Proceedings of the Annual Simulation Symposium*; Society for Computer Simulation International: Guildford/Surrey, UK, 2018; Volume 12, pp. 1–12. [\[CrossRef\]](#)
58. Barbati, M.; Bruno, G.; Genovese, A. Applications of agent-based models for optimization problems: A literature review. *Expert Syst. Appl.* **2012**, *39*, 6020–6028. [\[CrossRef\]](#)
59. Seber, G.A.; Lee, A.J. *Linear Regression Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2012; p. 329.
60. Rehman, K.U.; Khan, M.N.A. The foremost guidelines for achieving higher ranking in search results through search engine optimization. *Int. J. Adv. Sci. Technol.* **2013**, *52*, 101–110.
61. Ghasemaghaei, M.; Ebrahimi, S.; Hassanein, K. Data analytics competency for improving firm decision making performance. *J. Strateg. Inf. Syst.* **2018**, *27*, 101–113. [\[CrossRef\]](#)
62. Wedel, M.; Kannan, P.K. Marketing analytics for data-rich environments. *J. Mark.* **2016**, *80*, 97–121. [\[CrossRef\]](#)
63. Hair, J.F. Knowledge creation in marketing: The role of predictive analytics. *Eur. Bus. Rev.* **2007**, *19*, 303–315. [\[CrossRef\]](#)
64. Eurostat Statistics Explained. Total General Government Expenditure on Recreation, Culture and Religion, 2018 (% of GDP). Available online: [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Total_general_government_expenditure_on_recreation,_culture_and_religion,_2018_\(%25_of_GDP\)_png](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Total_general_government_expenditure_on_recreation,_culture_and_religion,_2018_(%25_of_GDP)_png) (accessed on 23 March 2020).
65. Grignard, A.; Taillandier, P.; Gaudou, B.; Vo, D.A.; Huynh, N.Q.; Drogoul, A.G. 1.6: Advancing the art of complex agent-based modeling and simulation. In *Proceedings of the International Conference on Principles and Practice of Multi-Agent Systems*, Dunedin, New Zealand, 1–6 December 2013; pp. 117–131. [\[CrossRef\]](#)
66. Sarlis, A.S.; Drivas, I.C.; Sakas, D.P. Implementation and dynamic simulation modeling of search engine optimization processes. Improvement of website ranking. In *Strategic Innovative Marketing*; Springer: Cham, Switzerland, 2017; pp. 437–443. [\[CrossRef\]](#)
67. Drivas, I.C.; Sakas, D.P.; Reklitis, P. Improving Website Usability and Traffic Based on Users Perceptions and Suggestions—A User-Centered Digital Marketing Approach. In *International Conference on Strategic Innovative Marketing*; Springer: Cham, Switzerland, 2017; pp. 255–266. [\[CrossRef\]](#)

