TOWARD INCREASING AND INVESTIGATING E-TOURISM DATA WAREHOUSE THROUGH A WEBSITES ANALYSIS STRATEGY

S. BOUREKKADI[1,2], K. SLIMANI[2], O. EL IMRANI[3], M.EZZAKI[2], A.BABOUNIA[2], Y.FAKHRI[2]

[1] Poitiers University, France

[2]Ibn Tofail University, Kenitra Morocco

E-mail: [2]khadija.slimani@uit.ac.ma , [3]oelimrani@uae.ac.ma

https://orcid.org/0000-0001-5264-2599
https://orcid.org/0000-0001-8036-2260
https://orcid.org/0000-0003-0080-0975

ABSTRACT

Nowadays, the competition between companies of all fields focuses on the information provided to the customers. International economic exchanges have experienced battles between large global companies because of the good marketing of their information. This guides us to confirm that the companies that perfectly manage their information related to their activities and know how to take advantage of the data that exist in their field are the companies that can keep an economic continuity. For this reason, the objective of this study is to work on an analysis strategy of websites especially for companies working in the field of tourism. This analysis will enable targeted companies to compete with their peers for informational competition to secure increased and investigated warehouses' data. This research is based on a multi-axis analysis as well as on a statistical study that collects the opinions of visitors on the used websites as case studies for this research. The main obtained results include the fact that several websites have many weaknesses. This is something that influences their performance. In addition, it was noticed that several websites' managers think that the performance of their product is based only on the number of visitors per day or the total of pages read in a specific period. In fact, these characteristics add nothing to the analysis when there are micro-conversions leading to macro-conversions.

Keywords: *Five Analysis, E-Tourism, Data Warehouse, Standby Strategy.*

1. INTRODUCTION

Day after day, new information and communication technologies become more important. They remain an effective and dynamic communication bridge when adapted with the latest types of technologies.

The use of these tools remains the best idea for a good performance. If we do not adapt them with standards, watch analysis, and audit strategy, we will not achieve a good performance .For that reason, companies must track such models of control and also track all its tools of communication with their targets especially the website, which manages and collects a large mass of data. With the growth of information, the company can ensure the data warehouses and ensure its increase with the days. For this to happen, our search can fill some gaps such as the unacceptable time of loading of the websites' contents, the lack of customer satisfaction on the management of the content (Websites), and other details that influence the performance of the Websites.

To attain its objectives, this article offers analysis and audit tools that allow to cite all the problems of the desired websites to analyze with lists of improvement tips. The list of problems in the previous table is a sampling. There is a lot of work going ahead to improve the content in order to avoid the maximum of these problems. That way, the loading speed of the website improves and the performance increases.

1. Unreachable resources
2. The PNG format is not the most suitable for few images.
3. Detection of some critical dependence.
4. Images are resized on the browser side.
5. The page should be accessible with and without 'www.' at the beginning of the URL.

6   Minimizing some JavaScript files (minification).
7   Reducing the number of DOM elements.
8   Specifying a set of characters in the HTTP headers response.
9   The addition of the ALT attribute to <img> tags
10 Some cookies are not secure
11 Several elements use the same identifier
12 Labels are not linked to an item
13 Some pages are subject to CLICKJACKING-type attacks

2.   RESEARCH METHODOLOGY

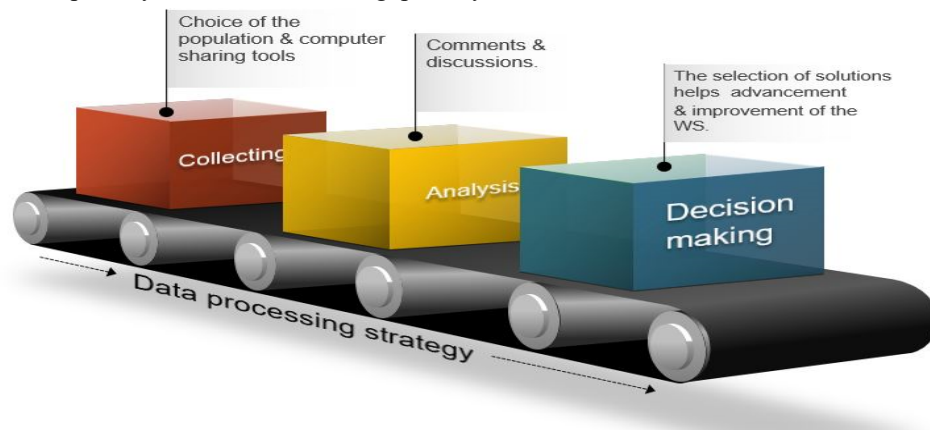As part of our survey, the questionnaire remains the primary means of collecting primary data. The questionnaire was used to obtain primary data on people's views, attitudes, opinions, and activities or behaviors. According to studies conducted by BLUEHOST, 90% of the users leave a site that takes more than 10 seconds to load and 60% leave sites that take more than 4 seconds. To ensure this, and also to ensure that visitors easily visit the websites of their requested service, a strategy of analysis of websites visitors has become necessary. Over time, this will ensure a data warehouse capable of helping managers to make strategic decisions and present their services in the best condition in a world where competition has become dependent on information. In the coming parts, the Main Steps of Applied Analysis are presented.

Figure 1: Data processing strategy

2.1 Data collection

The first stage of analysis starts with a data collection. Two questionnaires were created and proposed. The first one was dedicated for the evaluation of the website. It contains 12 questions. The second questionnaire was for the overall analysis of the website. It was made of 4 sections and consists of 15 questions. The questions that were chosen were transparent to a varying degree so that interviewees know the purpose of the research. The types of the used questions were as follows:

- Closed ended.
- Open ended.
- Open indirect.
- Closed indirect.

These questions were applied on two websites, which have the same purpose and all of which are in common (Tourist activity, presentation, and marketing of the same tourist city, the proposal of the same tourist service of the same city ...). In order to avoid promoting both websites, a decision has been taken so as to make the websites ambiguous and anonymous. The first website was provided with the WS1 symbol while the second competing website was different using a WS2 symbol. These symbols will be used in the rest of the research in order to come out with a comparison between these two websites (WS1 and WS2).

The objective of this collection of opinions is to give value, legitimacy, and credibility to the research especially as the participants were PhD and master researchers. They were from computer studies fields and information systems management.
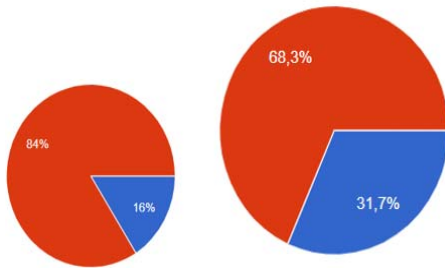
Figure 2: Percentage of participants by gender

Figure 2 demonstrates the percentage of the participants by gender in both surveys, evaluation survey (left) and analysis survey (right). Males participants are represented using the blue color while female participants are represented using the red color.

2.2 Technical analysis

In fact, the technical analysis can make the difference through the reference of the website, correcting its features, and ensuring the satisfaction of visitors (by ensuring a powerful search engine, contact forms, registration, visibility, and absence of pop-ups).

According to the analysis survey (the technical section), the following opinions of the participants were collected.

▪ Is the contact form clearly visible?

Regarding the visibility of the contact form, it was found that 56.7% of the participants were able to find it, 15% also succeeded but with some difficulties. Meanwhile, 28% confirm that this form is not visible.

▪ Is the possibility of registering clearly visible?

Regarding the visibility of the registration form, results show that 46.7% of the participants were able to find the registration form. Additionally, 26.7% succeeded but with some difficulties. However, 26.7% confirm that this form is not visible.
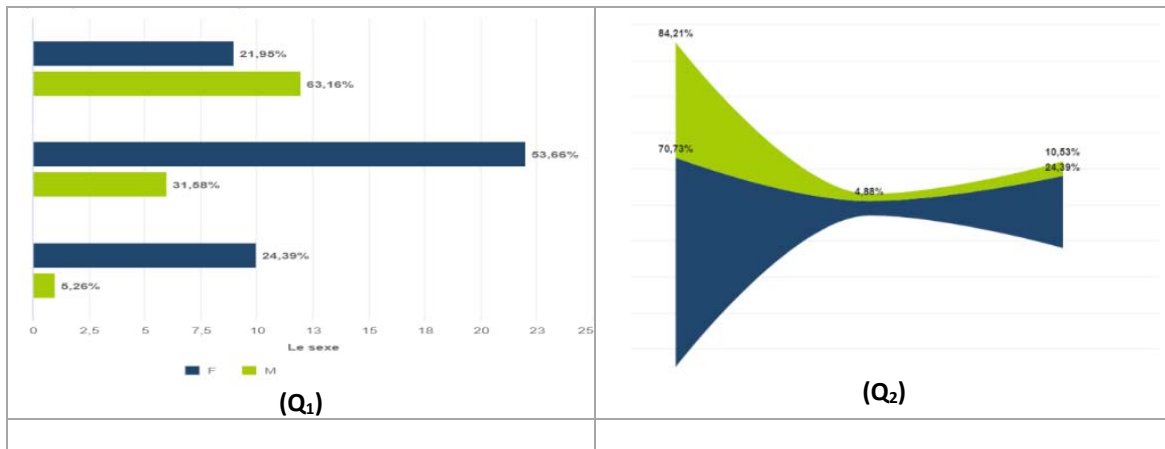
3. RESULTS

3.1. Content analysis

The content of a website is the whole text, images, and videos mentioned above. For this, it must be presented in a clear manner. From this content, the user must know the owner of the website, the page that tackles this issue. The home page should also present the sector of activity and the service/s supplied.

The content must also show the owner well with an easy-to-understand loading and this should spread over all the pages of the website.

The content analysis brings back the relations between the titles of the pages. This is to say for instance, finding out if there is coherence or not between the mentioned elements without neglecting the repetition of information and finding out whether it exists or not to do not approach the problem of duplicate or plagiarized content.
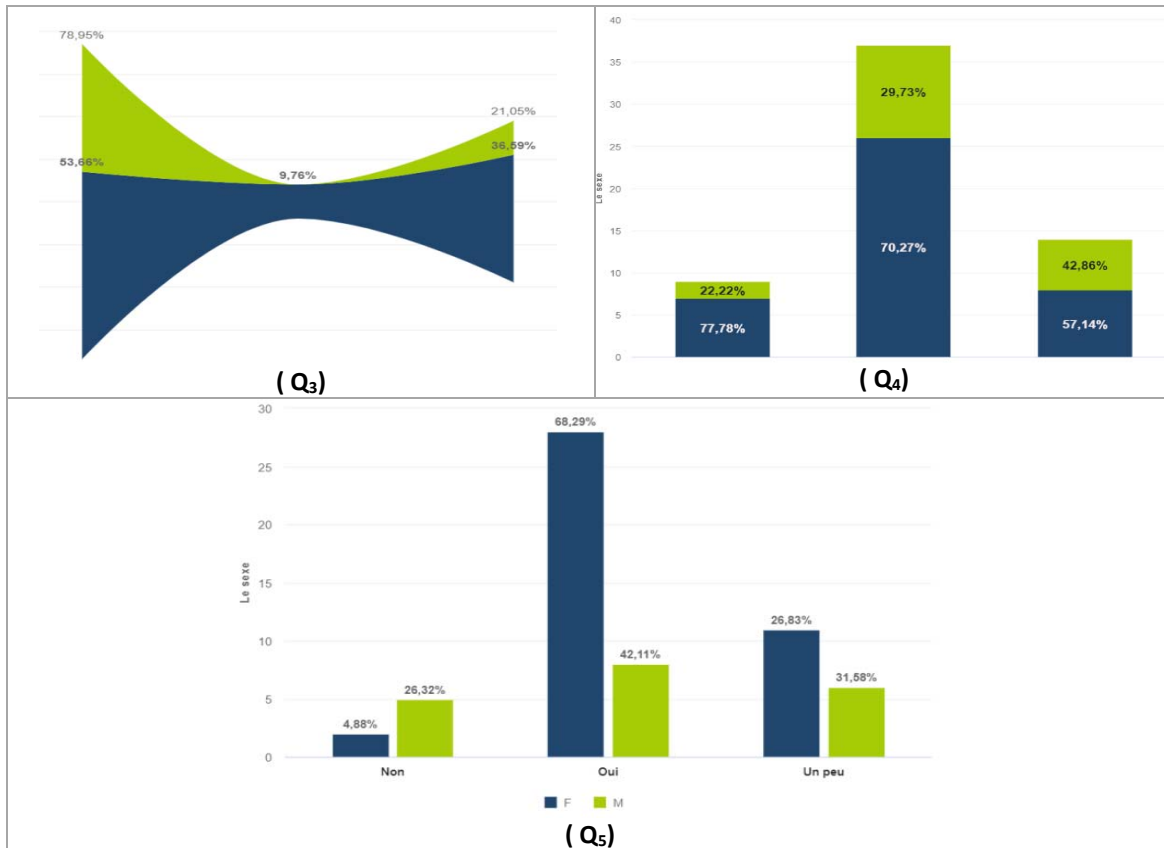

(Q₁)

(Q₂)

Figure 3: The 5 Q allowed the performance of a website

Q1-Did you understand who we are from your visit to the website?

Regarding the IDENTIFICATION OF THE OWNER OF THE WEBSITE, results show that 46.7% of the participants have succeeded to find it. Additionally, 18.3% of the participants have succeeded but with difficulties. However, 35% did not show a positive result.

Q2-Are there mistakes and typos on the website?

Regarding the complaints about spelling mistakes on the website, more than 75% of the participants mentioned a positive response, 5% have a negative response; however, 20% their response varies between positive and negative.

Q3-Are there duplicated contents?

Regarding duplicate contents on the website, more than 61.7% of the answers confirm the absence of plagiarism. The rest, 38.3% have a different answer 'between' yes and 'a little'.

**Q4-Does the Website adopt a tone that matches its target?**

For the relationship between the content of the website and its target, more than 61.7% of the participants agree so much that this relationship is well respected while 23.3% are more or less with the 61.7%. On the other hand, the rest, which presents 15%, is so much against the 85%.

Q5-Does the website highlight the advantages it offers to visitors?

Question 5 shows that 60% of the participants thinks that the website highlights the advantages it offers to visitors. On the contrary, 11.7% said no and 28.3% said "A little".

For checking duplicate content, it is possible to use several tools. Among these tools, Siteliner and Copyscape, both give acceptable results. The following figures provide a brief summary.
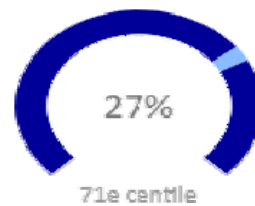
The duplicate content of the WS1 is 27%. The median for all other sites is 20%. Duplicate content on your website represents over 71% of all other sites.

On the WS1, 27% of the content is duplicated, 40% of the content is in common, and 32% of the content is unique.

With these tools too from the other results that were displayed, the following links are examples:

- https://www.sylvainluc.fr/ calculeo-isolation-a-1-euro-par-travaux-isolation-pour-1-euro/
- https://www.moulindelasalle.com/category/manifestations/avril-2020/
- https://m.vendee.cci.fr/actualite/visitez-la-criee-des-sables-dolonne
- https://www.lereportersablais.com/ vendee-focus-sur-les-frequentations-des-musees-vendeens
- https://www.lessables.mobi/fiche/detail/2773/ Visite-guid%C3%A9e-de-la-cri%C3%A9e-et-du-port-de-p%C3%AAche
- https://www.destination-lessablesdolonne.co.uk/Ticket-Service/ Booking-Office-voyages-concerts-spectacles
- https://www.ouest-france.fr/pays-de-la-loire/les-sables-dolonne-85100/ pays-des-olonnes-pensez-vous-inscrire-pour-les-visites-d-entreprise-6024686
- http://www.visiteznosentreprises.com/wp-content/uploads/2018/09/ Programme-JRVE2018-BD.pdf

3.2. Ergonomic analysis

The ergonomic analysis allows studying a set of axes to ensure if it is easy to navigate the Website. Among the objectives of this type of analyses:

- Ensuring access to the website pages from the main page with a few clicks;
- Securing a small menu that includes all the important pages;
- Securing a Map site;
- Securing a breadcrumb trail;
- Securinga guide button on the home page;
- Ensuring the accessibility, functionality, and compatibility of the website on all types of device and screen size;
- Securing a quick and logical loading of the website as 90% of internet users leave a website that takes more than 10 seconds to load and 60% for loadings of more than 4 seconds (study conducted by BLUEHOST);
- Ensuring optimal loading time;
- Ensuring a consistent presentation of the content of the website (Colors, Logos, Images, pictograms,...);
- Securing tools for the ergonomic audit of the website.

According to the analysis of the surveys in the ergonomics section, the following opinions were collected from participants (see Table 3 below).
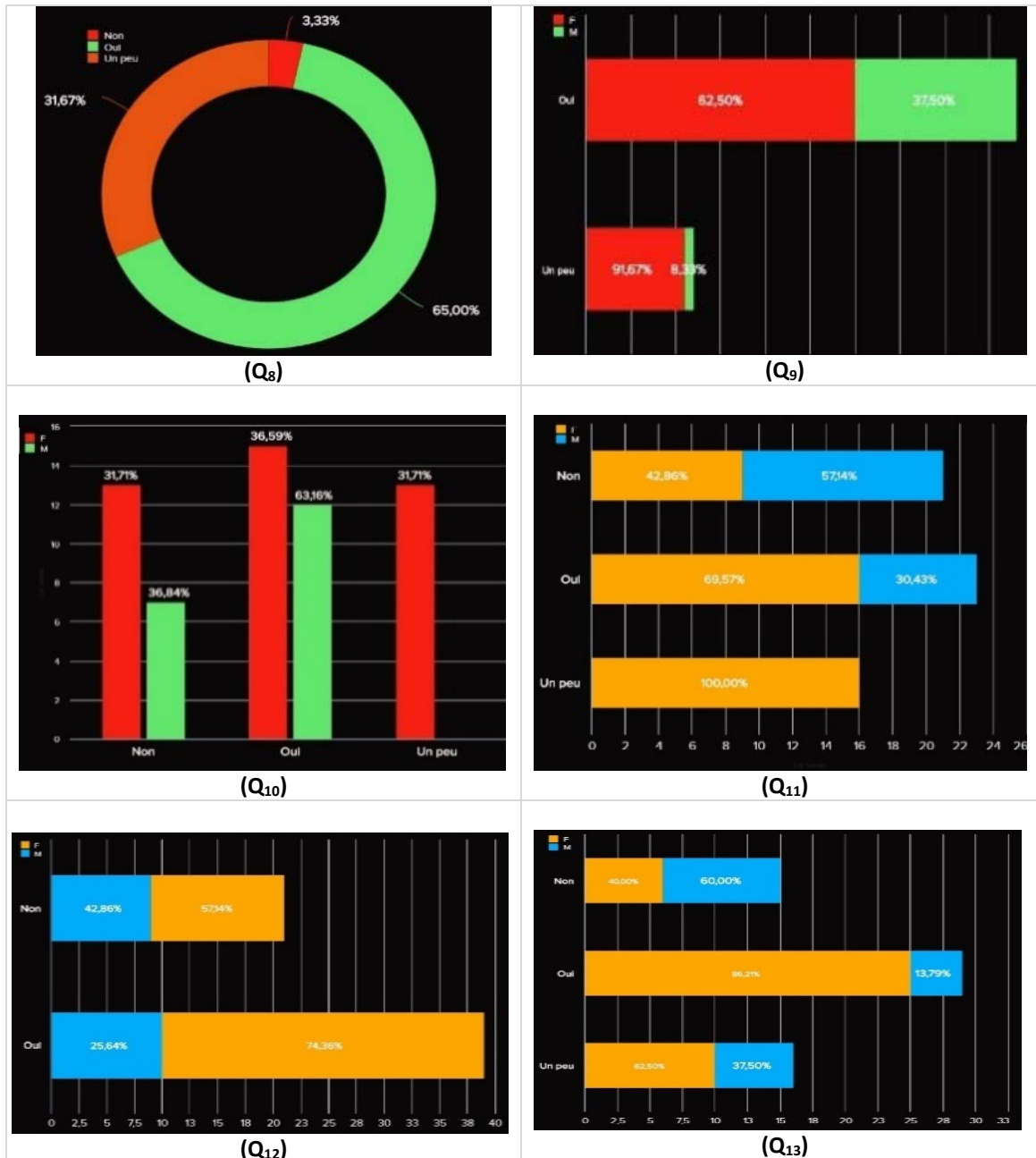

(Q6)

(Q7)
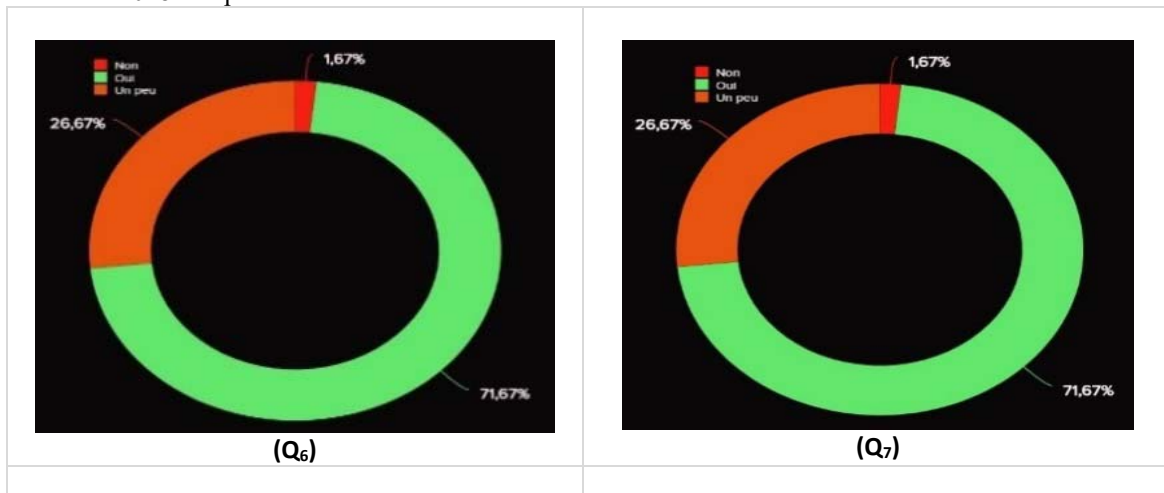
Figure 4: Ergonomic analysis results for Q6-Q13

Q6- Is the Website pleasing to the eye?
Results of question 6 show that 71.67% of the participants think that the website is pleasing to the eye. On the contrary, just 1.67% said the opposite and 26.67% said "A little".

Q7- Is everything on the Website harmonious?
The same results of question 6 are repeated at the level of question 7. This demonstrates that the website exceeded ¾ of positive answers about its form.

Q8- Is everything coherent from one end to the other?
For the consistency and coherence of the website, 65% of the participants confirm that the coherence of the website is well respected. In addition, 31.76% see that it is a little respected while only 3.33% have different views.

Q9- Do the pictograms help to understand the text they occur with?
For the role of pictograms on the website, 80% confirm that these pictograms help understanding the text they go with.

Q10- Do you get lost on the website?
Regarding this question, the results show that more than 45% of the participants said Yes, 33.3% said "No" whereas 21.7% said "a little".

Q11- Is it easy to navigate the Website?
Regarding the degree of difficulty of navigation on the website, the findings demonstrate that 38.3% of the participants have managed to navigate easily. In addition, 26.7% have succeeded to easily navigate but with some difficulties. However, 35% mentioned that they did not easily arrive to navigate the website.

Q12- Can you access all pages from the home page in 2 clicks maximum?
Regarding the access to all pages from the home page in 2 clicks maximum, 35% of the participants said that they have succeeded. On the contrary, 35% have not succeeded.

Q13- Does the website manage to guide you to the right place?
Regarding the orientation of browsers to visitors' objective of the website, it should be noted that 48.3% of the participants succeeded, 26.7% succeeded too but with some difficulties, and 25% failed to navigate to the right place.

From these results, tools of ergonomic analysis were collected. Although there are many of these tools, only two types were selected in our present article:
- Page Speed Insights.
- WEBPAGETEST.

The first website score is 10 on Mobile and 64 on the computer. Meanwhile, the weaknesses of the website are mentioned below with suggested examples of the correction in order to optimize the loading of web pages.
- Consider loading hidden or off-screen images after that all essential resources have been loaded to reduce the delay before interactivity.
- Optimized images load faster and consume less mobile data.
- Minimizing JavaScript files can reduce the size of payloads and the time it takes to scan scripts.
- Minimizing JavaScript files can reduce the size of payloads and the time it takes to scan scripts.
- (…)

A further analysis of the results, which will allow the passage to the end comb of any movement and which happens on the website1 (WS1), we can arrive at the procedure of loading our tested website. Time is necessary for a virtual visitor to complete the loading, see figures 5.
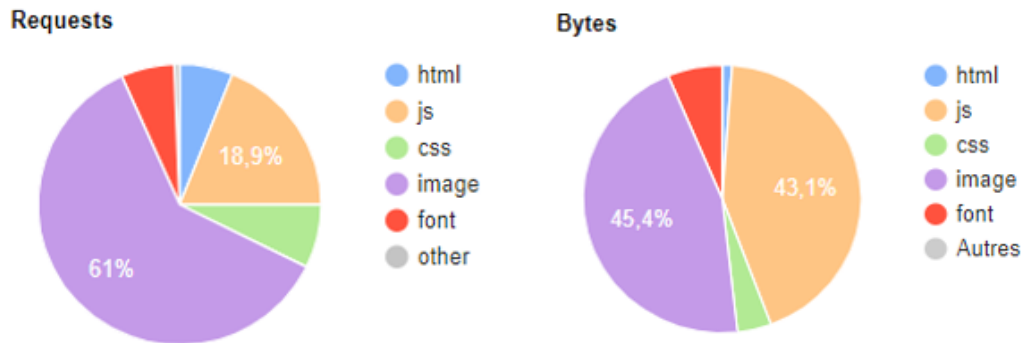

Figure 5: loading information WS1

The implementation of these correction proposals can give estimates of savings of more than 11.72 seconds of loading time on mobile and more than 3.65 seconds on computer. This indicates that the speed of this example page is slow.

3.3. Performance Analysis
The analysis of the performance of the website is not only based on the number of visitors per day or the total of the pages read in a specific period. These values do not almost add anything to the analysis. Sometimes it is important to ask some questions to get a clear idea about the performance of the website. For instance by asking the following questions:
- How often a video is viewed on the website?
- Is there a space for collecting visitors' reviews?
- Is there the possibility that a customer can comment on articles on the website?
- How many visitors have visited the contact page?

- Is there a reaction with forms on the website?

It is essential also to ask if there is an analysis of the collected information because with small details it is possible to achieve great success.

As an example, the collected results of an important question are suggested below.

Q$_{16}$- Have you seen a video of the website?

Figure 6: 4WS1 audience overview
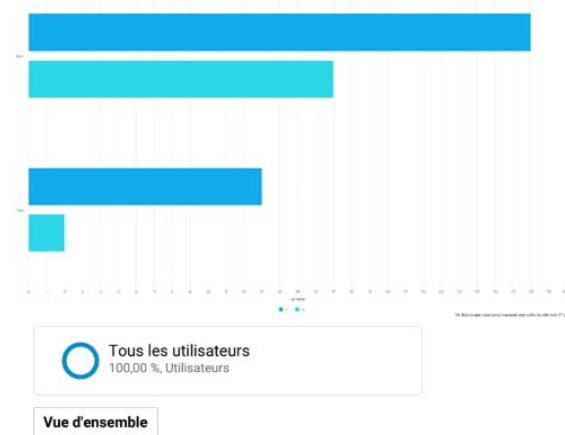
For the last question, have you watched a video of the website? Results show that 75% of the participants answered with No. On the contrary, just 25% answered with Yes.

In the figure, the female participants are represented in blue, and the masculine in soft green.

To strengthen performance analysis, Google analytics tool was used. The latter is used to track the micro and macro conversions. It is sufficient to create the objectives that come in a few types:

- A user accesses a given page.
- The user spends some time on the website.
- The user reads a number of pages.
- The user performs an action.

3.4. Technological Analysis

The objective of the technological analysisis to get the technology with which a website is developed (Programming language and Content management system). For that reason, the following figure shows the most used CMS according to IONOS (data from February 2020)
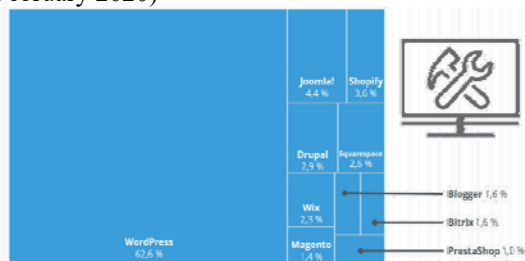

Figure 7: CMS Statistics

3.5. Word positioning - the right vocabulary-

In a website, it is necessary to position the keywords well in order to make them visible when typing a query on a search engine. For this purpose,

the experts of the field confirm that it is very profitable to invest more time in the positioning of your website by choosing the right keywords; those that guide you to ensure a good positioning in a web-marketing audit.

Analyzing the position of your keywords

First, check the list of keywords currently used, and try to improve it by other keywords that you find important and interesting; those that allow you to easily find your website on search engines.Second, ensure for each set of keywords a list of pages they are positioned on. The list can give you a clear idea about SEO. We suggest the following tools to carry out the process:

- *Google Search Console* (Webmaster Tools)
  o Knows the words used to access the website.
  o Knows the total display of the website (impressions).
  o Knows the number of clicks of a user on your page.
  o More services: Google Help page.

- *SEObserver* "SEO control panel"
  o Adds positioning keywords.
  o The participating pages in the first ranking to your positioning.
  o Analysis of competitors' keywords.
- *Yooda* SEO Benchmark
  o Improves your keyword spreadsheet.
  o Displays the positions of the website in the Google search engine.
  o Displays a list of competitors' keywords.
- *Yooda* Keyword Suggest
  o Suggests keywords related to your domain, based on a keyword or keywords list.

4. DISCUSSION

After analyzing and comparing the WS1 and WS2, the following results have been achieved:

Weight by type of resource –Ko-
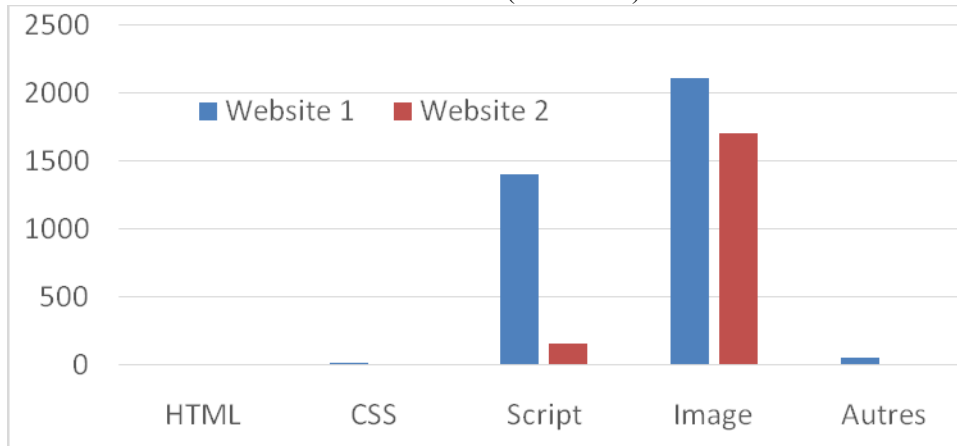For HTML and CSS, WS1 and WS2 almost have the same KB. On the other hand, there is a big difference between the SCRIPTs and images (see figure 8).

Number of requests by the type of resource
At the level of the number of requests by the type of resource, the WS1 has a higher number of requests than that of WS2. This number is sometimes tripled especially at the level of SCRIPT, image and others images (see figure 9).

Loading time
For the loading time, the weaknesses of WS1 are always observed in all levels (First octet, Start of display, End of visual display).Besides, the end of loading of WS1 almost doubles that of WS2(see figure 10 for more information on the speed index) images (see figure 10).

Visual progress
The list of improvements in the previous table is also a sampling of a set of proposals. In fact, working on the implementation of these tips can yield many advantages and positive consequences (See table 1).

Figure 8: Weight by type of resource –Ko

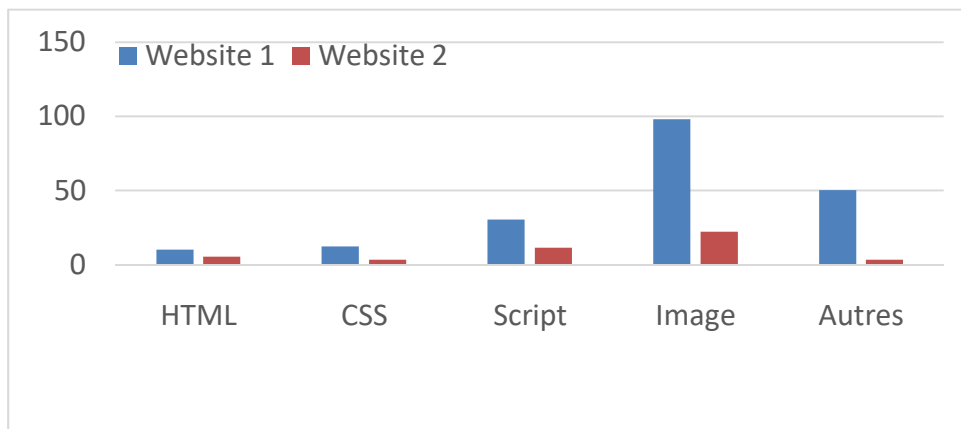*Figure 9: Number of requests by the type of resource*
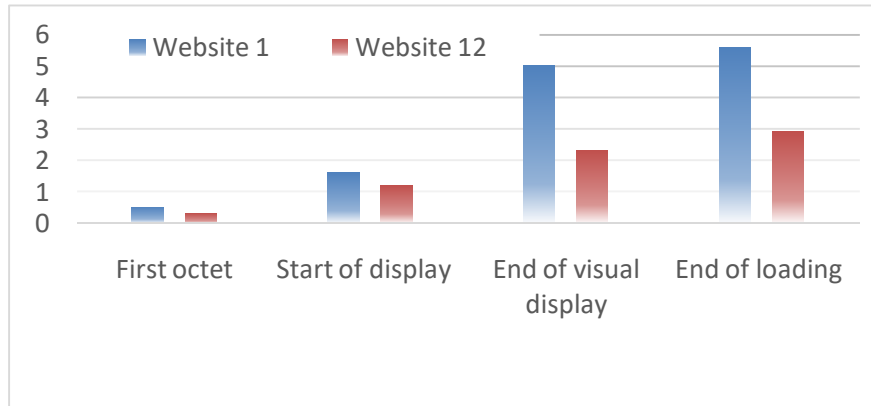
Figure 10: Loading time

Figure 11: Speed index

Table 1: List of the Improvements (WS1)

|	List of the Improvements
1	Requests do not define a cache policy.
2	Images are quickly loaded.
3	Empty items can disrupt screen readers.
4	Some resources do not define the type of content
5	Some iframes can be secured with the SANDBOX attribute
6	Some CSS properties are overloaded by SHORTHAND
7	Some CSS selectors are duplicated
8	Explanation of the purpose of each form field.
9	The keyword '!important' is used several times
10	Third-party resources are delivered without integrity control (SRI)
11	No use of INLINE scripts that are too long
12	Separation of HTML styles
13	Improvement of site reading by social networks
14	Some CSS properties are duplicated
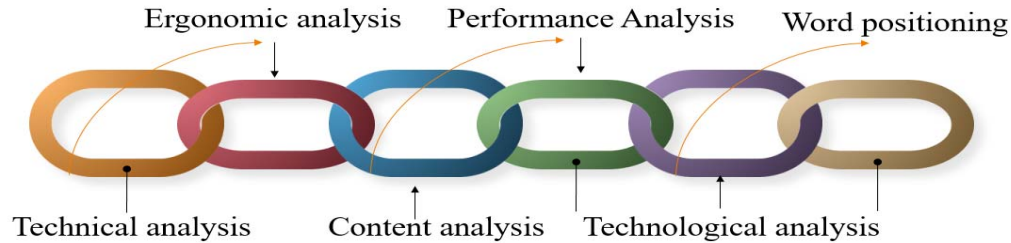
Figure 12: Website Analysis Architecture

5. CONCLUSION

Based on the findings, it can be concluded that the analysis of the websites requires steps to follow. In addition, ensuring the collection of data on our websites helps offering a complete and comprehensive analysis of the targeted website. Once mini-data are collected, day after day, the data warehouse can be provided to help with ideal decision-making in order to ensure logical and efficient competition.

On the internet network, SEO is literally vital. Try asking the right question about the importance of a website if no one gives it an interest. Also, properly manage the positioning of a website at the base of its keywords is a difficult mission, which requires a lot of research. The keywords must be chosen wisely since that helps getting the chance to be consulted by the targeted internet users.

REFRENCES:

[1]. Gauthier, B. (2003). *Recherche sociale: de la problématique à la collecte des données*. Puq.

[2]. Antoine, P., Bry, X., & Diouf, P. D. (1987). La fiche AGEVEN: un outil pour la collecte des données rétrospectives. *Techniques d'enquête*, *13*(2), 173-181.

[3]. Lelièvre, É. (1999). Collecter des données de mobilité: des histoires migratoires aux biographies d'entourage. *Espace populations sociétés*, *17*(2), 195-205.

[4]. Miles, M. B., & Huberman, A. M. (2003). *Analyse des données qualitatives*. De Boeck Supérieur.

[5]. Meulman, J. J. (1992). The integration of multidimensional scaling and multivariate analysis with optimal transformations. *Psychometrika*, *57*(4), 539-565.

[6]. Wicaksono, B. L., Susanto, A., & Winarno, W. W. (2012). Evaluasi Kualitas Layanan Website Pusdiklat BPK RI Menggunakan Metode Webqual Modifikasian dan Importance Performance Analysis. *MEDIA*, *19*(1).

[7]. Shia, B. C., Chen, M., & Ramdansyah, A. D. (2016). Measuring customer satisfaction toward localization website by WebQual and importance performance analysis (case study on AliexPress Site in Indonesia). American Journal of Industrial and Business Management, 6(2), 117-128.

[8]. Himawan, A. K. (2014). Performance Analysis Framework Codeigniter and CakePHP in Website Creation. International Journal of Computer Applications, 94(20).

[9]. Yoric Mercuzot, Audit de site web : les 4 aspects d'un site à analyser pour évaluer son potentiel, -easybear-

[10]. Malo, N., & Warren, J. (2011). *Web Analytics: Mesurer le succès et maximiser les profits de votre site web*. Editions Eyrolles.

[11]. Bonaccorsi, J. (2013). Approches sémiologiques du web. *Manuel d'analyse du web en Sciences Humaines et Sociales*, 125-146.

[12]. Charrad, M., Lechevallier, Y., Saporta, G., & Ahmed, M. B. (2008). Web Content Data Mining: la classification croisée pour l'analyse textuelle d'un site Web. In *EGC* (pp. 43-54).

[13]. MOKRANE, A., AREZKI, R., DRAY, G., & PONCELET, P. (2004, March). Cartographie automatique du contenu d'un corpus de documents textuels. In *Proceeding of the 7th international conference on the statistical analysis of textual data JADT, 12-15 mars 2004, Louvain-La-Neuve, Belgique.*.

[14]. kaothar-meziane 'step by step website audit, journal ducm 2020.15797

[15]. Yoric Mercuzot "Website audit: the 4 aspects of a site to analyze to assess its potential', 2020. – Easy Bear -