Clustering Analysis of Usability in Web Sites of Higher Technological Institutes of Ecuador

Yeferson Torres-Berru[1](✉) and Pablo Torres-Carrión[2](✉)

[1] Instituto Superior Tecnológico Loja, Av. Granada y Turunuma, 1101608 Loja, Ecuador
ymtorres@tecnologicoloja.edu.ec
[2] Universidad Técnica Particular de Loja, San Cayetano Alto S/N, 1101608 Loja, Ecuador
pvtorres@utpl.edu.ec

Abstract. Techniques for evaluating usability continue to be innovated. This document shares the application of a heuristic-based framework for measuring web usability - SIRIUS, complemented by two machine learning techniques for clustering: a) Hierarchical, with the Ward.2 method and Euclidean; and b) K-means clustering. For data processing, CRISP-DM has been proposed as a general method. Since our objective is to evaluate the usability characteristics of the websites of the Technical and Technological Institutes of Ecuador, data has been obtained from the web portals of 83 Institutes (34 public and 49 private). As a result, three clusters have been obtained, which encompass the 10 aspects of the framework, and which allow us to identify the levels of usability of technological institutes. As a result, 18 institutes have been categorized into the group of websites with above-average usability (cluster1), 32 institutes with below-average usability (cluster2), and the remainder with an acceptable degree of usability. The method used and proposed has made it possible to have a general usability map of the web portals of the technical and technological institutes of a country, as input for decision-making.

Keywords: Clustering · Heuristics · SIRIUS · Usability · Unsupervised learning

1 Introduction

The Higher Education System of Ecuador, to date there are 59 accredited universities, of which three only offer postgraduate degrees, and 83 Technical and Technological Institutes, being 34 public and 49 private[1]. Each institution has a public web domain, from which academic and management resources are administered. After having carried out a Systematic Literature Review, no studies on the usability of all institutional websites have been found. In [1] n analysis is made of 24 websites of universities belonging to CEDIA, using the Prometheus tool from SIRIUS and Kmeans, obtaining as relevant data that 50% are at an adequate level, but fail to highlight in which aspects they present problems. In [2] the websites based on the ISO 9241-151 standard in 59 universities are analyzed, classifying them by the category obtained in the evaluation by CACES (A,

[1] https://www.educacionsuperior.gob.ec/ Sitio Web de la Secretaria de Educación Superior, Ciencia, Tecnología e Innovación del Ecuador.

V. Agredo-Delgado et al. (Eds.): HCI-COLLAB 2020, CCIS 1334, pp. 63–72, 2020.
https://doi.org/10.1007/978-3-030-66919-5_7

B, C), also detecting problems in content design, general design and especially in the search criteria. In [3] the study at the University of Riobamba is singled out, determining that the main problems are the use of Search Engine and Interacting with Links and Download Information. This research complements the previous ones, completing the entire Ecuadorian higher education system. The proposed objectives are of interest to higher education authorities, knowing the usability of their web spaces, which are the point of contact with the community.

To assess usability on the web, there are many methods and techniques, which, Perurena [4] very accurately details. Granollers [5] proposes the use of heuristics, as a technique to evaluate usability in user interfaces; Storm [6] uses time Series Analysis of Selected Episode Graphs; In [7] they make use of the ISO/IEC 9126 quality model standard, applying techniques such as observation guide, satisfaction questionnaire and attractive interface questionnaire to evaluate each metric. In [8] card sorting method is applied to construct the shape of the site focused on user shared cognition. In [9] the use of the eye tracking tool is exposed, as an input for the analysis of user behavior when interacting with the web. Artificial intelligence has also been applied to enhance data collection and perform sustained analysis in machine learning algorithms; Thus, Zainab [10] applies clustering techniques as an input to categorize data obtained from the web, which are then analyzed from a standard system usability scale. To improve the automation of software evaluation (including usability), Stouky [11] proposes the use of machine learning techniques and big data. Chamba [1] applies data mining techniques to collect data from 24 Universities Web portals, as input for the subsequent application of usability techniques. In this work, Chamba's proposal [1] is followed to collect data from 83 institutional websites.

In this research, it is proposed to work with the User-Oriented Web Usability Assessment System based on the determination of Cryptic Tasks (SIRIUS by its Spanish acronym), developed by Torrente [12] as part of his work on PhD research; It is composed of 10 aspects, divided into 83 variables (criteria) that cover the various areas of functionality of a web-based solution. Torrente exposes it in [13] like as a heuristic-based framework for measuring web usability adapted to the type of website. Sai Aparna [14] adapts it as SIRIUS-DWUEP (Web Usability Evaluation Process), operationalizing this empirical validation technique in the WebML method. Muñoz-Egido [15] applies it to analyze eighteen academic library web portals, obtaining as a result that the average usability of the websites of the analyzed university libraries is 72.30 out of 100 with a standard deviation of 5.57. In [1] his technique is used in 24 Universities Web portals, obtaining three clusters that relate the 10 aspects of SIRIUS. In this research, SIRIUS is applied, and techniques based on clustering that have allowed us to categorize the web portals of the 63 institutes of higher education since the fulfillment of each of the aspects of this technique.

2 Materials y Methods

As a general method of automatic learning, the CRISP-DM method is proposed [16], adapted to the singular requirements of the study. This method is composed of six iterative stages: business understanding, data understanding, data preparation, modeling, evaluation and deployment.

2.1 Business Understanding

The usability evaluation of 83 public (34 eq = 40%) and private (49 EQ = 60%) institutes websites was performed following the Sirius framework: A heuristic-based framework for measuring web usability adapted to the type of website [13]. Sirius has 83 evaluation criteria grouped into ten aspects, which are detailed in Table 1. The data for the analysis was obtained through a project carried out at the "Instituto Tecnológico Loja", and is available on the web[2], in open access format.

2.2 Data Understanding

2.3 Data Modeling

Two clustering techniques are used for data modeling:

a) **Hierarchical**, with the Ward.2 method and euclidean distance between two points p and q, which is defined as the length of the segment that joins both points. Hierarchical combined with descriptive stadia [17], o determine the suitable number of centers for Kmeans. Once the dendrogram is created, the extent to which its structure reflects the original distances between observations is evaluated, using the correlation coefficient between the cophenetic distances of the dendrogram (node height) and the original distance matrix.
b) **K-means** clustering [18] to group the observations in K different clusters (centers), based on the evaluations previously carried out, it was determined that the value of K = 3, with 100 iterations to guarantee the process.

The computational analysis was carried out with the R Studio[3] software, using the packages: *clValid* to validate the ideal clustering algorithm; *factoextra* for Multivariate Analysis); and, *cluster* and *stats* for Clustering Analysis; In addition to the official sites, these packages can be downloaded from the project website[4]. To find the similarity, the Euclidean distance is used, making a hierarchical classification where the groups are merged (or subdivided) successively, following a priority or hierarchy, decreasing the homogeneity as they become wider. Therefore, for the evaluation of clustering, the coefficient of correlation is considered, which allows finding the distance between the clusters.

[2] https://bit.ly/UsabilityInstitutes.
[3] https://rstudio.com/.
[4] https://torresyeferson.github.io/UsabilityInstitutes/.

Table 1. Sirius aspects and criteria's

Aspect	Criteria
General Aspects (GA)	*10 variables*: goals of the site are concrete and well defined, contents and services are precise and complete, general structure of the site is user-oriented, general look & feel is aligned to the goals of the website, general design of the website is recognizable, general design of the website is coherent, user's language is used, other languages are supported, translation of the page is complete and correct, website is updated regularly
Identity and Information (II)	*7 variables*: identity or logo is significant, identifiable and visible, identity of the website is present in every page, slogan or tagline is suited to the goal of the site, information about the website or company is provided, contact mechanisms are provided, information about privacy of personal data and copyright of web contents is provided, information about authorship, sources, creation and revision dates of articles, news and reports is provided
Structure and Navigation (SN)	*14 variables*: welcome screen is avoided, structure and navigation are adequate, element organization is consistent with conventions, number of elements and terms per element is controlled in navigation menus, depth and breadth are balanced in the case of hierarchical structrure, links are easily recognized as such, link depiction indicates its state, redundant links are avoided, broken links are avoided, self-links to the current page are avoided, image links indicate the content to be accessed, a link to the beginning of the page is always present, elements hinting where the user is and how to undo the navigation exist
Labelling	*6 variables*: labels are significant, labelling system is precise and consistent, page titles are planned and correct, home page url is correct, clear, and easy to remember, inner page urls are clear, inner page urls are permanent
Layout of the page	*10 variables*: higher visual hierarchy areas of the page are used for relevant content, information overload is avoided, clean interface with no visual noise, white areas between information objects are provided for visual rest, visual space on the page is used correctly, visual hierarchy is correctly used to express "part of" relationships between page elements, page length is under control, print version of the page is correct, page text can be read easily, zblinking/moving text is avoided
Comprehen-sibility and ease of interaction	*7 variables*: concise and clear language is used, language is user friendly, each paragraph expresses an idea, interface controls are used consistently, visible metaphors are recognizable and comprehensible by any user, coherent or alphabetic order indrop-down menus, available options in a user-input field can be selected instead of written
Control and feedback	*10 variables*: user controls the whole interface, user is informed about what is happening, user is informed about what has happened, validation systems are in place to avoid errors before the user sends information, clear and non-alarmist information, and recovery actions are provided to the user when an error has occurred, response time is under control, website windows cancelling or superimposing over browser windows are avoided, proliferation of windows is avoided, user downloading of additional plugins is avoided, in task with several steps, user is informed of the current step and the number of steps remaining to complete the task
Multimedia elements	*6 variables*: images are well-cropped, images are comprehensible, images have the correct resolution, some added value is provided by using images or animations, cyclical animations are avoided, some added value is provided by using sound

(*continued*)

Table 1. (*continued*)

Aspect	Criteria
Search	*8 variables*: if necessary is accessible in every page, easily recognizable, easily accessible, text box width is enough, simple and clear search system, advanced search is provided, search results are comprehensible for the user, user is assisted in case of empty results for a given query
Help	*5 variables*: help link is located in a visible and standard place, easy access to and return from the help system, context help is offered for complex tasks, faq query selection and redaction is correct, faq answers are correct
10 aspects	*83 variables (criterias)*

3 Results

In Fig. 1 we observe the grouping of the institutes applying the Hierarchical method and obtaining 3 groups with a correlation coefficient of 0.9526396.

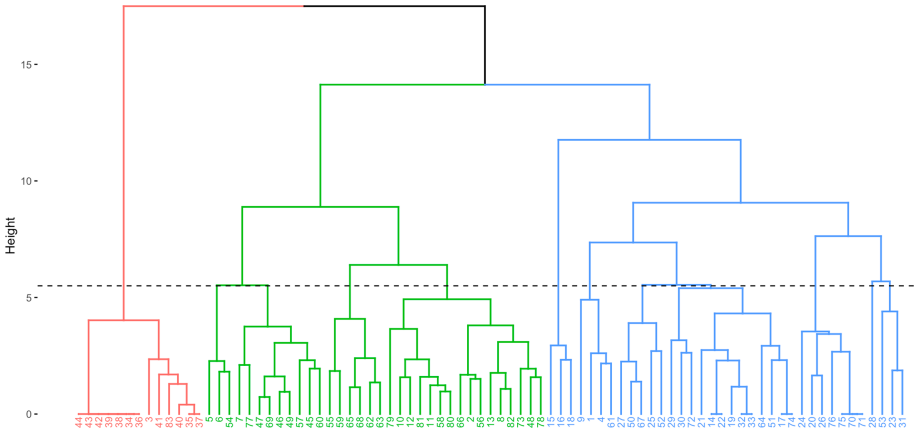

Fig. 1. Hierarchical cluster

The clustering analysis allowed to identify 3 groups of institutes (Fig. 2) with 18, 32 and 33 institutions respectively. Regarding the classification by financing of the institutions, we find that Cluster 1 is made up of 17 private and one public institute, Cluster 2 of 17 private and 15 public, and cluster 3 of 25 private and 8 public.

In Fig. 3 the values of the cluster with respect to the 10 aspects of SIRIUS are shown. Cluster1 predominates in 2 aspects (search and help); Cluster 2 predominates in 4 aspects (General aspects, layout, control and multimedia); and cluster3 predominates in 4 aspects (identity and information, structure and navigation, labeling and comprehensibility).

Table 2 details the distribution of the 3 general clusters and the clustering of the 83 criteria among the 10 aspects evaluated by SIRIUS, classifying the institutes between public and private. These significance values allow us to observe that 73.49% of the institutes evaluated obtained a high evaluation in Labeling, besides that only 25.50%

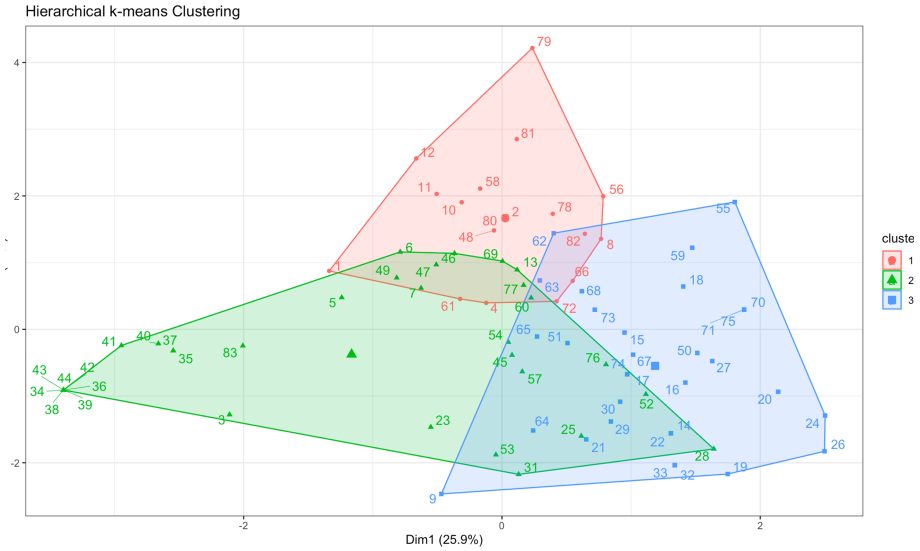

Fig. 2. Kmeans cluster

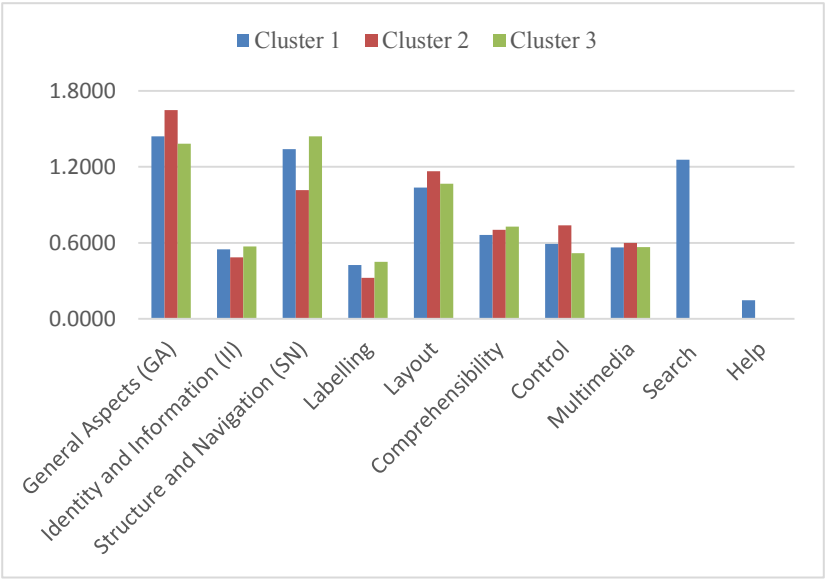

Fig. 3. Classification aspects

obtained a high evaluation in Control, we also highlight that 88% of institutes do not have a search incorporated in the website, finally we observed that only 19% of institutes show user help on site.

Table 2. Cluster Sirius Aspects

GA	II	SN	Labelling	Layout	Comprehensibility	Control	Multimedia	Search	Help
CLUSTER 1 (Usability Average 80.04 min_value = 66 max_value = 84)									
1.4417	0.5497	1.3404	0.4256	1.0369	0.6622	0.5911	0.5651	1.2561	0.1467
n = 34	n = 30	n = 45	n = 43	n = 46	n = 15	n = 33	n = 43	n = 10	n = 13
priv = 24	priv = 20	priv = 35	priv = 36	priv = 30	priv = 13	priv = 21	priv = 32	priv = 2	priv = 9
pub = 9	pub = 10	pub = 10	pub = 7	pub = 16	pub = 12	pub = 12	pub = 11	pub = 8	pub = 4
CLUSTER 2 (Usability Average 64.71 min_value = 54 max_value = 72)									
1.6491	0.4863	1.0170	0.3233	1.1652	0.7030	0.7397	0.5990	0.00	0.0000
n = 43	n = 16	n = 17	n = 22	n = 23	n = 41	n = 22	n = 1		n = 67
priv = 9	priv = 8	priv = 8	priv = 6	priv = 18	priv = 32	priv = 12	priv = 0		priv = 47
pub 15	pub = 8	pub = 9	pub = 16	pub = 5	pub = 9	pub = 10	pub = 1		pub = 20
CLUSTER 3 (Usability Average 71.25 min_value = 71 max_value = 82)									
1.3825	0.5713	1.4419	0.4503	1.0670	0.7283	0.5199	0.5657	0.000	0.0017
n = 6	n = 37	n = 21	n = 18	n = 14	n = 17	n = 28	n = 39		n = 3
priv = 6	priv = 31	priv = 16	priv = 17	priv = 11	priv = 14	priv = 26	priv = 27		priv = 3
pub = 0	pub = 36	pub = 5	pub = 1	pub = 3	pub = 3	pub = 2	pub = 12		pub = 0

Finally, to better understand the usability margins of the three clusters, a graph with measures of central tendency and a usability center line are detailed in Fig. 4. It is observed that the percentage of usability of the institutes belonging to the cluster1 is above the average; however, the institutes belonging to cluster2 and cluster3 are below the average, with better results for the last group.

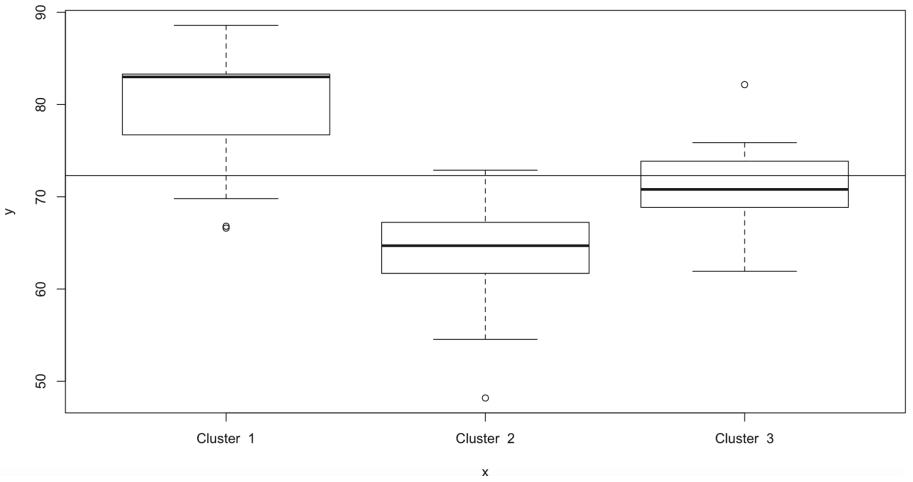

Fig. 4. BoxPlot cluster – usability percent

4 Discussion

Being our objective to evaluate the usability characteristics of the institutes' websites, we have identified 3 types of institutes and the most significant aspects are search and help, the search aspect presents similarity with problems previously found in university websites [1, 3]. As indicated by research [19, 20] the user must be able to retrieve information easily, therefore, it is important to provide adequate search systems, because the websites of educational institutions allow the presentation of academic offerings, news and achievements in teaching, research and networking. In this area, the importance of the understanding of a website is highlighted [21, 22], as shown in this research, most Ecuadorian institute sites present information, but care must be taken to ensure that this is done through appropriate web interfaces that speed up understanding.

The analysis based on heuristics used in this research, and adaptable to the type of website, can be adapted to the information and usability needs of educational institutions; this presents a similarity to the work [2], However, research such as [3] use the ISO 9241-151 standard, which evaluates all types of sites with the same metrics; furthermore, these studies are based on the categorization of type A, B, C universities etc. and the present study, classifies according to the analysis performed by the artificial intelligence technique used, based on the 81 parameters and 10 aspects evaluated with SIRIUS.

Of the aspects found with SIRIUS, the one corresponding to Help is shown as a weakness in most institutes, so it presents an opportunity to decongest the traditional communication channels and provide the user with a better web experience in educational sites [23], being this an aspect to improve, implementing sections of frequent questions, tutorials, chatbots etc.

5 Conclusions

Group 1 is characterized by high values in Search and Help and low values in Understandability and Multimedia elements, Group 2 is characterized by high values in Page Layout, Control and Feedback, low values in Tagging and null values in Search and Help; on the other hand Group 3 presents high values in Site Identity, Structure and Navigation, Tagging, Understandability, Multimedia elements, low values in Help and null values in Search.

Evaluating the aspects and criteria of usability of 83 technological institutes of Ecuador we can conclude that Cluster 1 composed by 18 institutes is a group that to improve its usability must improve its understandability and multimedia elements, Cluster 2 composed by 32 institutes is a group whose scores must improve in most aspects except in Page Layout and Control and Feedback, on the other hand Cluster 3 composed by 33 institutes is a group whose values are high in 6 of the 10 but must improve in help and implement search in the web site.

Acknowledgement. We thank the students of the Interface Design Subject of the Instituto Superior Tecnológico Loja Extension Vilcabamba for the support in the data collection.

References

1. Chamba-Eras, L., Jacome-Galarza, L., Guaman-Quinche, R., Coronel-Romero, E., Labanda-Jaramillo, M.: Analysis of usability of universities web portals using the Prometheus tool - SIRIUS. In: 2017 4th International Conference on eDemocracy eGovernment, ICEDEG 2017, pp. 195–199 (2017). https://doi.org/10.1109/ICEDEG.2017.7962533
2. Pincay-Ponce, J., Caicedo-Ávila, V., Herrera-Tapia, J., Delgado-Muentes, W., Delgado-Franco, P.: Usabilidad en sitios web oficiales de las universidades del Ecuador. Rev. Ibérica Sist. Tecnol. Inform. (2020)
3. Rosas-Chavez, P., Mora-Fernandez, J., Suarez, C.: Comparative analysis of usability of the public universities' web sites of Riobamba City in Ecuador. In: Ahram, T., Falcão, C. (eds.) AHFE 2019. AISC, vol. 972, pp. 742–752. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-19135-1_73
4. Perurena Cancio, L., Moráguez Bergues, M.: Usability of Web sites, methods and evaluation techniques. Rev. Cuba. Inf. Cien. Salud. **24**, 176–194 (2013)
5. Granollers, T.: Usability evaluation with heuristics. New proposal from integrating two trusted sources. In: Marcus, A., Wang, W. (eds.) DUXU 2018. LNCS, vol. 10918, pp. 396–405. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91797-9_28
6. Storm, K., Kraemer, E., Aurrecoeche, C., Heiges, M., Pennington, C., Kissinger, J.C.: Web site evolution: usability evaluation using time series analysis of selected episode graphs (2009). https://doi.org/10.1109/WSE.2009.5630633
7. Salazar-Grandes, M.C., et al.: Usability evaluation mechanism with standard ISO/IEC 9126. Case study: Tourism portals (2017)
8. Shieh, J.-C., Lin, H.-W.: The study of web findability based on its breadth and depth. J. Educ. Media Libr. Sci. **50**, 255–288 (2013). https://doi.org/10.6120/JoEMLS.2012.502/0484.RS.CM
9. Menges, R., Kramer, S., Hill, S., Nisslmueller, M., Kumar, C., Staab, S.: A visualization tool for eye tracking data analysis in the web (2020). https://doi.org/10.1145/3379156.3391831
10. Zainab, S.S.E., Mehmood, Q., Zehra, D., Dietrich, R.-S., Hasnain, A.: PrEVIEw: clustering and visualising pubmed using visual interface (2016)
11. Stouky, A., Jaoujane, B., Daoudi, R., Chaoui, H.: Improving software automation testing using jenkins, and machine learning under big data. In: Jung, J.J., Kim, P., Choi, K.N. (eds.) BDTA 2017. LNICSSITE, vol. 248, pp. 87–96. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-98752-1_10
12. Suárez Torrente, M.D.C.: Sirius: sistema de evaluación de la usabilidad web orientado al usuario y basado en la determinación de tareas críticas (2011)
13. Torrente, M.C.S., Prieto, A.B.M., Gutiérrez, D.A., De Sagastegui, M.E.A.: Sirius: a heuristic-based framework for measuring web usability adapted to the type of website. J. Syst. Softw. **86**, 649–663 (2013). https://doi.org/10.1016/j.jss.2012.10.049
14. Sai Aparna, S., Baseer, K.K.: SIRIUS-DWUEP: a heuristic-based framework for measuring and evaluating web usability in model-driven web development (2015). https://doi.org/10.1007/978-3-319-13728-5_34
15. Muñoz-Egido, D., Osti, M.V: Evaluation of usability of Spanish academic library web portals using a cognitive-emotional heuristic model. Rev. Esp. Doc. Cient. **40** (2017). https://doi.org/10.3989/redc.2017.1.1379
16. Wirth, R., Hipp, J.: CRISP-DM: towards a standard process model for data mining. In: Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, pp. 29–39. Springer, London (2000)
17. Hotelling, H.: A generalized T test and measure of multivariate dispersion. In: Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 23–41 (1951)

18. Macqueen, J.: Some methods for classification and analysis of multivariate observations (1967)
19. Martín Fernández, F.J., Iazza, G., Hassan, Y.: Diseño web centrado en el usuario: usabilidad y arquitectura de la información - hipertext - (UPF), pp. 1–15 (2014)
20. Folmer, E., Bosch, J.: Architecting for usability: a survey. J. Syst. Softw. **70**, 61–78 (2004). https://doi.org/10.1016/S0164-1212(02)00159-0
21. Yusof, U.K., Khaw, L.K., Ch'ng, H.Y., Neow, B.J.: Balancing between usability and aesthetics of Web design. In: Proceedings 2010 International Symposium on Information Technology - Visual Informatics, ITSim 2010 (2010). https://doi.org/10.1109/ITSIM.2010.5561310
22. Sindhuja, P.N., Dastidar, S.: Impact of the factors influencing website usability on user satisfaction. IUP J. Manag. Res. **8**, 54–66 (2009)
23. Calisir, F., Bayraktaroğlu, A.E., Gumussoy, C.A., Topcu, Y.I., Mutlu, T.: The relative importance of usability and functionality factors for online auction and shopping web sites. Online Inf. Rev. **34**, 420–439 (2010). https://doi.org/10.1108/14684521011037025