Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset of measurable indicators of the Slovak state institutions' websites

Veronika Bojdová*, Viliam Ďuriš, Ľubomír Rybanský

Department of Mathematics, Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra, Tr. A. Hlinku 1, 949 01 Nitra, Slovakia

ARTICLE INFO

ABSTRACT

The dataset contains 11 measurable indicators for the website evaluation from different points of view. These indicators were collected for 60 websites of the Slovak state institutions. It provides information about the directly measurable variables, which may affect or reflect the usability, popularity and visibility of the website. Most variables were measured by online tools. The dataset is a mixture of binary, ordinal, discrete numeric and continuous numeric variables, which gives many opportunities to analyze the relations between the measurable websites' indicators. It can be used to find the structure consisting of latent variables, which cannot be directly measured (such as usability or popularity of the website). Another use is to find subgroups of state institutions, which have similar websites from some point of view.

© 2024 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

---

* Corresponding author.
  E-mail address: vbojdova@ukf.sk (V. Bojdová).

Specifications Table

Subject	Computer Science: Human-Computer Interaction
Specific subject area	Measurable indicators of the Slovak state institutions' websites
Type of data	Table, Raw
Data collection	Directly measurable indicators of the Slovak state institutions' websites were collected by using online tools:
	• https://www.seotesteronline.com • https://www.deadlinkchecker.com • http://validator.w3.org • http://tools.pingdom.com • https://search.google.com/test/mobile-friendly • https://www.rankwatch.com/free-tools/alexa-rank-checker • https://www.backlinkwatch.com
	Some websites did not allow measuring of some variables, therefore missing values coded by "NA" are present in the data.
Data source location	Department of Mathematics, Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra, Tr. A. Hlinku 1, 949 01 Nitra, Slovakia
Data accessibility	Repository name: ZENODO Data identification number: 10.5281/zenodo.10560140 Direct URL to data: https://doi.org/10.5281/zenodo.10560140

1. Value of the Data

- The data focuses on the websites of the Slovak state institutions and consists of measurable evaluation indicators. It provides information about the directly measurable variables, which may affect or reflect the usability, popularity and visibility of the website.
- The dataset [10] is a mixture of binary, ordinal, discrete numeric and continuous numeric variables, which gives many opportunities to analyze the relations between the measurable websites' indicators.
- Data can be used for Structural Equation Modeling or Partial Least Squares Path Modeling to find the structure of the inner model consisting of latent variables (such as usability or popularity of the website), which cannot be directly measured. Another use is to find subgroups of state institutions, which have similar websites from some point of view (Latent Profile Analysis, Cluster Analysis).

2. Background

The motivation to create this dataset was the question of whether the websites of Slovak state institutions have similar features or parameters that can be measured in some way. Are the websites of all state institutions approximately the same, do they look similar, or are they divided into different subgroups of similar websites? It was necessary to determine which indicators affect the usability, popularity, and findability of the website and can be measured directly with some online tool. A methodology for the evaluation of websites of Slovak small and medium-sized enterprises proposed in [1] dealt with similar questions, and was helpful to choose proper indicators to measure. We focused on some general concepts, e.g., popularity, functionality or usability of the websites, which are reflected in directly measurable variables (indicators) connected to them. Selected indicators (Table 1) were chosen according to the known website evaluation approaches (see [2]).

Table 1

An overview of the raw data table.

Column name	Description	Variable type	Missing data
web.ID	unique anonymized identifier of the website	character	0
SEO	SEO rating from the online SEO Scoring Service (https://www.seotesteronline.com was used), with a range from 0 (worst) to 100 (best), this indicator measures the visibility of the website.	numeric	0
links.broken	ratio of non-functional links to the total number of links of the given website (https://www.deadlinkchecker.com was used). This indicator influences the functionality of the website.	numeric	4
errors	number of errors on the homepage (HTML validator (http://validator.w3.org was used), this indicator is a consequence (effect) of the website's functionality.	numeric	0
speed.grade	speed of the website was measured by the Pingdom Website Speed Test (http://tools.pingdom.com was used). To achieve objectivity, the Google Page Speed performance grade was used, with a range from 0 (worst) to 100 (best). This indicator is a consequence (effect) of the website's functionality.	numeric	0
correct.view	value 1 = website is suitable for mobile devices, value 0 = website is not suitable for mobile devices. This indicator influences the usability of the website (https://search.google.com/test/mobile-friendly was used).	binary	0
clear.navigation	value 0 = website has neither navigation nor map, value 1 = website has either navigation or a map, value 2 = website has both navigation and a map. This indicator influences the usability of the website.	ordinal	0
language.versions	number of language versions of the website. This indicator influences the usability of the website.	numeric	0
Alexa.rank	an estimate of the website's popularity measured by Alexa Traffic Rank (https://www.rankwatch.com/free-tools/alexa-rank-checker was used). Numbers greater than 1 million are replaced by the symbol ">1M". This indicator measures the popularity of the website.	numeric/text	0
relative.Alexa	recalculated indicator from Alexa.rank, as the ratio of the Alexa.rank to the 1 million. This indicator reflects the popularity of the website - the smaller the number, the more popular the website.	numeric	0
referring.domains	number of referring domains which refer to the website (https://www.backlinkwatch.com was used). The greater the number, the better. This indicator measures the visibility of the website.	numeric	0
world.ranking	an estimate of the website's popularity measured by World Ranking (https://www.backlinkwatch.com was used). The smaller the number, the better. This indicator is an effect of the website's popularity.	numeric	0

3. Data Description

The dataset is an Excel file with 2 sheets: the first sheet contains raw data and the second sheet contains the description of the variables. Each row of the raw data presents one website (Table 1) from a total of $N = 60$ cases.

Fig. 1 presents barplots for categorical variables in the dataset. Table 2 summarizes main characteristics of numeric variables and Fig. 2 provides boxplots and histograms for numeric variables in the dataset.

Numeric variables in the dataset are uncorrelated (Fig. 3, no significant correlations marked), that is an advantage for using them in statistical methods with the assumption of non-correlated input variables.
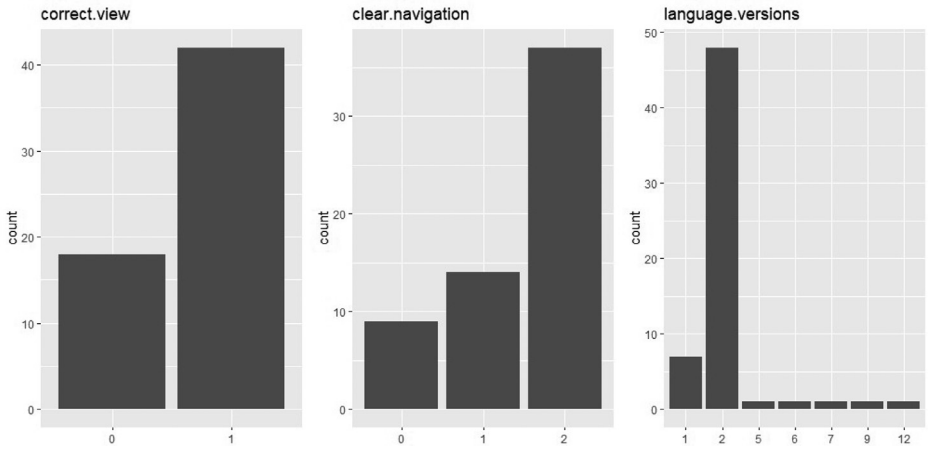
Fig. 1. Barplots for categorical variables in the dataset.

Table 2
Summary characteristics for numeric variables in the dataset.

|	SEO	links.broken	errors	speed.grade	relative.Alexa	referring.domains	world.ranking
Min.	50.000	0.000	0.000	66.000	0.006	2.000	0
1st Quartile	60.000	0.007	5.000	70.750	0.821	33.750	99,255
Median	67.000	0.017	12.500	75.500	1.000	77.500	1,192,820
Mean	66.300	0.064	30.220	75.950	0.802	171.770	1,781,229
3rd Quartile	72.000	0.032	23.500	80.000	1.000	184.250	2,378,852
Max.	85.000	0.726	71.000	97.000	1.000	102.000	9,781,249
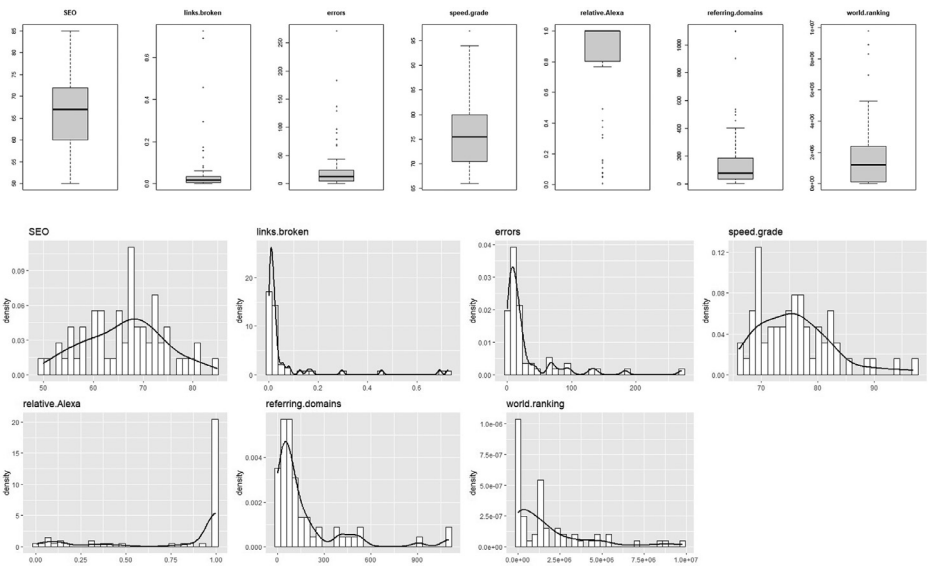
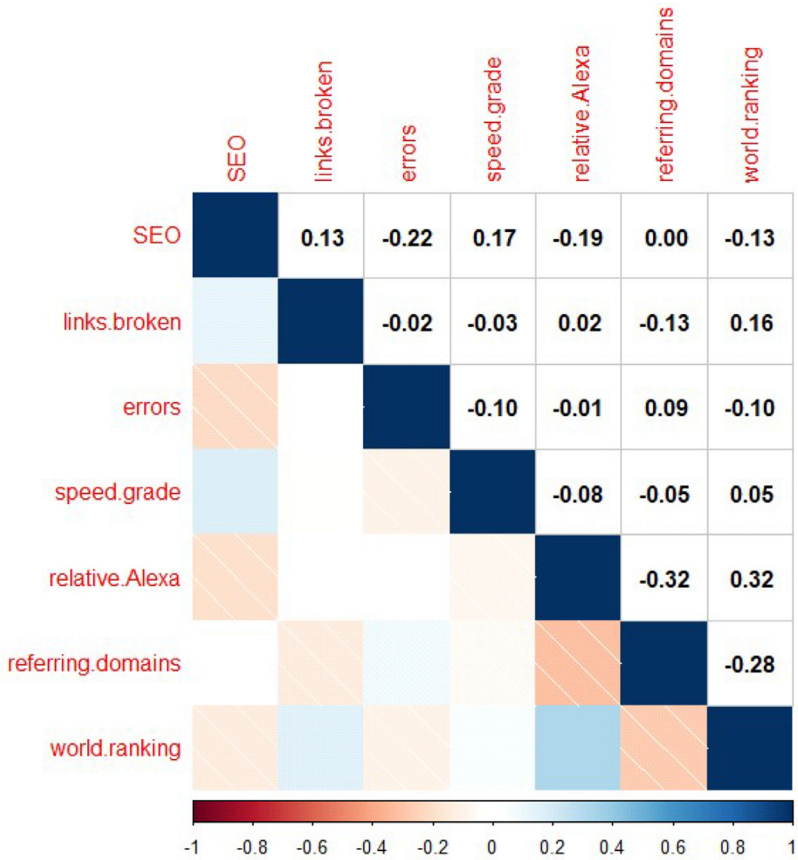**Fig. 2.** Boxplots and histograms for numeric variables in the dataset.

Fig. 3. Pearson correlation matrix of the numeric variables (significant correlations marked with∗).
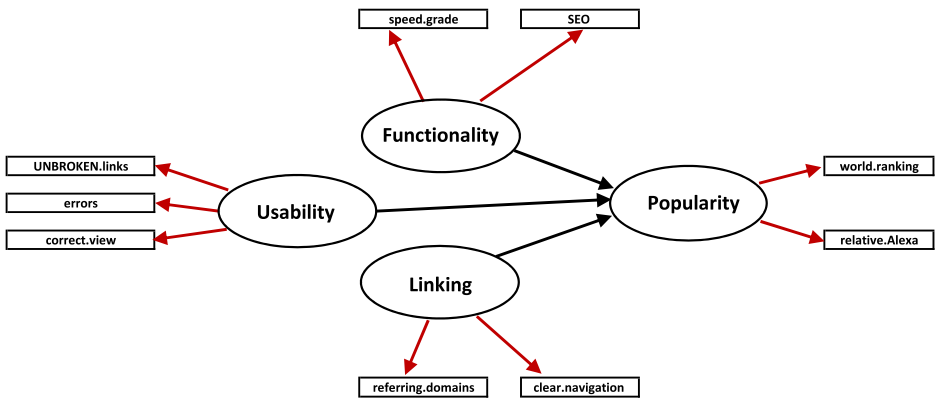

Fig. 4. An example of a Partial Least Squares Path Model with 4 latent variables Functionality, Usability, Linking and Popularity.
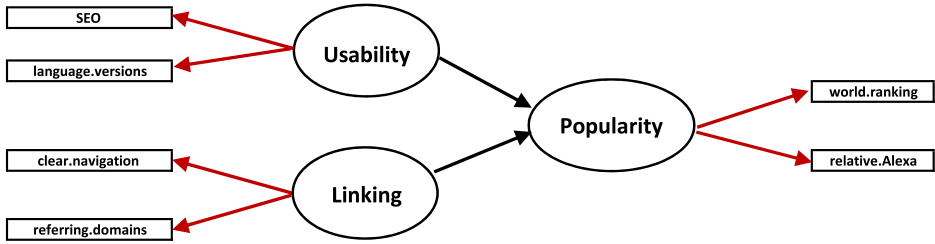
Fig. 5. An example of a model with 3 latent variables Usability, Linking and Popularity.

An example of the Partial Least Squares Path Model (Fig. 4) assumes, that the inner model consists of 4 latent variables

- Functionality – represents the website functionality, which reflects 2 directly measurable indicators **speed.grade** and **SEO**,
- Usability – represents the website usability, which reflects 3 directly measurable indicators **links.broken, errors** and **correct.view**,
- Linking – reflects 2 measurable indicators **clear.navigation** for the inner navigation structure of the website, and **referring.domains** for the outer links referring to that website,
- Popularity – represents the overall website popularity given by all previous latent variables, which reflects two directly measurable indicators **world.ranking** and **relative.Alexa**.

An example of the Partial Least Squares Path Model (Fig. 5) assumes, that the inner model consists of 3 latent variables

- Usability – reflects 2 directly measurable indicators **SEO** and **language.versions**,
- Linking – reflects 2 measurable indicators **clear.navigation** for the inner navigation structure of the website, and **referring.domains** for the outer links referring to that website,
- Popularity – represents the overall website popularity given by two previous latent variables, which reflects two directly measurable indicators **world.ranking** and **relative.Alexa**.

Fig. 6 presents the result of Latent Profile Analysis of the dataset, where two groups of websites were created with different profiles in the observed variables. Each group contains websites with similar profiles and different profiles to the other group.

4. Experimental Design, Materials and Methods

The data collection was performed from March to May 2023. The list of the Slovak state institution websites was created from the official government websites (https://www.minv.sk/?internetove-stranky-statnych-institucii, https://www.vlada.gov.sk/adresar-institucii/). The final set of 60 websites consist of all government websites and websites of the Slovak state institutions (e.g. National Bank of Slovakia, Financial Administration of the Slovak Republic). We did not make any sampling, we used all functional websites that allowed measuring their metrics by online tools (some websites has very strict security policy). We used only nationwide websites, not the regional or district institutions' webpages.

The following tools were used to gather the metrics (variables in Table 1) of the websites:

- https://www.seotesteronline.com [3] – for SEO rating (variable **SEO**), with a range from 0 (worst) to 100 (best),
- https://www.deadlinkchecker.com [4] – for the ratio of non-functional links to the total number of links of the given website (variable **links.broken**),
- http://validator.w3.org [5] – for the number of errors of the homepage (HTML validation, variable **errors**),
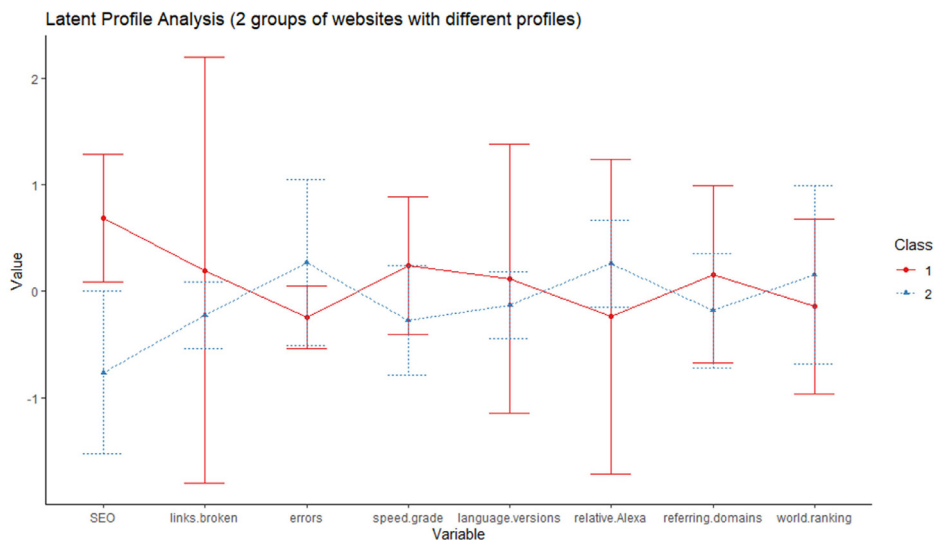
Fig. 6. An example of Latent Profile Analysis result.

- https://tools.pingdom.com [6] – the Pingdom Website Speed Test for measuring the speed of the website (variable **speed.grade**). To achieve objectivity, the Google Page Speed performance grade was used, with a range from 0 (worst) to 100 (best). In more detail described at http://developers.google.com/speed,
- https://search.google.com/test/mobile-friendly [7] – for binary variable **correct.view** indicating whether the website is suitable for mobile devices (value 1) or not (value 0),
- https://www.rankwatch.com/free-tools/alexa-rank-checker [8] – for an estimate of the website's popularity measured by Alexa Traffic Rank (variable **Alexa.rank**). It is calculated using a combination of average daily visitors and pageviews of specific websites over the past 3 months. The smaller the number, the more visited the website. The value of Alexa Traffic Rank is constantly changing for a given website over time. If the resulting numbers are greater than 1 million, the tool replaces them with the symbol ">1M".
- https://www.backlinkwatch.com [9] – for the number of referring domains that refer to the website (variable **referring.domains**) and for the estimate of the website's popularity measured by World Ranking (variable **world.ranking**).

The number of language versions of the website (variable **language.versions**) and values for ordinal variable **clear.navigation** were assigned by visual assessment of the website.

Because of a mixture of numeric and text data in the variable **Alexa.rank**, we decided to recalculate this indicator to have all values numeric. The recalculated indicator **relative.Alexa** is the ratio of the **Alexa.rank** to the 1 million. This way we get numeric values for all cases. Unpopular websites would have a ratio of 1 and most popular websites would have a ratio close to 0. This indicator reflects the popularity of the website in the same way as the original **Alexa.rank** variable - the smaller the number, the more popular is the website.

Finally, the dataset was anonymized by replacing the website links with the unique identifier of the website (variable **web.ID**).

Limitations

Online tool https://www.deadlinkchecker.com (used to obtain values of variable **links.broken**) checks only up to 2000 links of the website, more links are a paid service.

Therefore, by calculating the ratio of non-functional links to the total number of links of the given website, the maximum total number of links is always less than or equal to 2000.

Some websites did not enable access to the internal structure of the website (due to the very strict security policy), therefore the measurement of the variable **links.broken** by dead-linkchecker [4] was unsuccessful and four missing values coded by "NA" are present.

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief and confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

Measurable indicators of the Slovak state institutions' websites (Original data) (ZENODO).

CRediT Author Statement

Veronika Bojdová: Conceptualization, Visualization, Writing – original draft, Writing – review & editing; **Viliam Ďuriš:** Data curation, Writing – review & editing; **Ľubomír Rybanský:** Methodology, Supervision, Visualization, Writing – review & editing.

Acknowledgements

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. Kráľ, P. Laco, Corporate website evaluation using partial least squares path modelling, in: Proceedings of the 22nd International Scientific Conference on Applications of Mathematics and Statistics in Economics (AMSE 2019), Atlantis Press, 2019, pp. 91–99, doi:10.2991/amse-19.2019.10.
[2] Chiou C, C.Lin C, Perng, A strategic framework for website evaluation based on a review of the literature from 1995–2006, Inf. Manag. 47 (5–6) (2010) 282–290.
[3] Quarzio s.r.l.Seo Tester Online, 2023 [date of access Apr., 2023] https://www.seotesteronline.com.
[4] Jonathan Grove, D.L.C. Websites, Dead Link Checker, 2023 [date of access Apr., 2023] https://www.deadlinkchecker.com.
[5] W3CMarkup Validation Service, 2013 [date of access May 2023] https://validator.w3.org.
[6] SolarWinds Worldwide, LLCPingdom Website Speed Test, 2023 [date of access May 2023] https://tools.pingdom.com.
[7] Google LLCMobile Friendly Testing Tool, 2023 [date of access Mar., 2023] https://search.google.com/test/mobile-friendly.
[8] Web IQ Technologies Pvt. LtdRankWatch's Alexa Ranking Checker, 2023 [date of access June 2023] https://www.rankwatch.com/free-tools/alexa-rank-checker.
[9] Withheld for Privacy ehfBacklink Watch, 2023 [date of access Mar., 2023] https://www.backlinkwatch.com.
[10] V. Bojdova, V. Ďuriš, Measurable indicators of the Slovak state institutions' websites, Zenodo (2024), doi:10.5281/zenodo.10560140.