Analysis on Web Technologies Usage of the Macedonian Web

Antonio Gilev, Bisera Bojkovska, Jovana Gelebesheva,
Ana Markovska, Marko Andonod and Boban Joksimoski, PhD
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, Macedonia

antonio.gilev@students.finki.ukim.mk, bisera.bojkovska@students.finki.ukim.mk,
jovana.gelebesheva@students.finki.ukim.mk, ana.markovska.1@students.finki.ukim.mk,
marko.andonov@students.finki.ukim.mk, boban.joksimoski@finki.ukim.mk

Abstract—The analysis on web technologies usage of the Macedonian web is a study that investigates the different web technologies that have been used to build Macedonian websites. This study covers a wide range of web technologies, including PHP, JavaScript, jQuery and many others. The study uses the Wappalyzer tool to collect data from the 320 most visited websites in Macedonia and identify the technologies used by these sites. In general, the analysis of the use of web technologies on the Macedonian web provides valuable insights into which technologies are most prevalent, as well as how many of the sites use exotic and non-supported versions of technologies. This information can be useful for web developers, designers and other stakeholders in Macedonia who are interested in building effective and efficient web solutions.

Index Terms—web, technologies, trends, web-analysis

I. INTRODUCTION

Some say that you can predict the future by studying and analyzing the past and present. The Macedonian web presence constitutes a minor fraction of the vast expanse of the World Wide Web. Given that our country is rapidly transitioning to a more digital landscape, it may be beneficial to chart the evolution of the digital landscape in relation to global trends. This paper will conduct preliminary inquiries into the technologies used to build the Macedonian web space and attempt to correlate and map it accordingly. The paper is organized as follows: chapter II will elaborate on the need for such analysis and examine similar research; chapter III will describe the methods used for data acquisition and initial processing, while chapter IV will present our initial insights into the data. The conclusion (chapter V) summarizes our findings and provides guidelines for our further work.

II. BACKGROUND WORK

A survey investigated the use of web technologies on the official websites of 548 local governments in Indonesia. To collect data for local government websites, the authors of this research used Wappalyzer. The study found that websites were not built from scratch, but web developers used different frameworks to develop them. It also found a lack of web 3.0 implementation and recommended the implementation of schema.org to realize open government for local government in Indonesia [1].

Another study explored the top 1,500 Alexa-ranked sites for Indonesian visitors and identified the types of apps they use along with their versions. It found that more than half of the applications in use are no longer supported by their maintainer. This underscores the urgent need for website owners to update their applications with the latest supported versions to avoid potential security threats and other issues. Furthermore, the study cites WordPress and Apache as successful because they encourage users to update to a supported version [2].

Several researchers have also analyzed websites using automated tools. They used Site Analyzer and Qualidator to analyze and evaluate the websites of the University of Punjab. Websites are evaluated by Site Analyzer based on their content, design, performance, and optimization, while Qualidator evaluates them according to usability, accessibility, and optimization. The results showed that Site Analyzer ranked www.puchd.ac.in as the best site while Qualidator ranked www.thapar.edu as the best site [3].

The Wappalyzer tool also helped in XSS, CSRF and SQL injection analysis in Colombian software and website development. Wappalyzer has been used to classify websites based on their name, economic sector, programming language, CMS, and web framework. This study checks 130 Colombian websites for XSS, CSRF, and SQL injection vulnerabilities using the Acunetix Trial Edition tool. The selection of websites for evaluation was based on several criteria, including the main economic sectors in Colombia, leading companies within those sectors, the most frequently visited websites in Colombia according to Alexa rankings, and websites developed by Colombian software companies. After the analysis, many security vulnerabilities were discovered. The study concludes that many Colombian web developers have ignored protection mechanisms even though the basic techniques for attacking websites are well known. The authors call for a focus on web security training for developers and collaboration between universities and industry to bridge the gap between education and real-world skills. [4].

In this research, the identification of the technologies as well as their version is done automatically with the help of the Wappalyzer library.

III. METHODOLOGY

As part of the study, it was necessary to identify and collect the domains of the most frequently used websites in Macedonia (*data-gathering* phase). Subsequently, an analysis was performed on each domain to determine the technologies utilized (the *analysis* phase).

A. Gathering the website list

Since the shutdown of Alexa [5], the premier web-analysis company, indexing website popularity has become a non-trivial task. Hence, the process of data gathering involved conducting research in order to obtain information about the most-visited sites on the Macedonian web. In orded to gather more relevant data, specialized online platforms for business analysis were utilized. Popular alexa.com alternatives, including StatCounter, BuiltWith and SimilarWeb, were used to identify the majority of Macedonian websites.

Specific choices were made during the compilation of the list:

- Domains that end with *.mk* do not have to be Macedonian websites. The services were also queried regarding the presence of Macedonian and/or Albanian language in the website content.
- Popular services that provided Macedonian-language versions of their websites, but were not based in North Macedonia (e.g. *google.mk*) were excluded.
- Only top-level domains were taken into consideration. For example, one of the most popular forums, hosted at forum.kajgana.com was not taken into consideration, but kajgana.com, a general portal was part of the dataset.

Using this approach, approximately 320 website domains were collected and subsequently subjected to analysis. The aim was to target as many categories as possible, some of which are e-commerce, banks, the educational sector, online newspapers, blog sites, the governmental sector, and more. The final list was stored as a CSV (Comma Separated Values) file, consisting of URL links to those websites.

B. Data retrieval experimental setup

The subsequent step involves gathering a significant amount of information about each website. While various approaches exist for obtaining this information, we chose to rely on a well-known and highly regarded service Wappalyzer [6]. To fully automate the process, the Python programming language was utilized. A Python script was created that analyzes all the gathered website domains. The python-Wappalyzer library [7] allows for the identification of various technologies, such as used web servers, backend programming languages, javascript frameworks and libraries, CMS platforms on a given website.

Additionally, our script utilized several other well-known libraries, including requests, pandas, docx, matplotlib.pyplot, numpy, and plotly.graphobjs.

IV. DATA ANALYSIS

There are various properties that can be analyzed from both the front-end (client-side) and back-end (server-side) perspectives. The goal of this paper is to provide initial insights into the state and trends of Macedonian web pages by analyzing the data collected.

One of the more interesting parts is trying to disclose the preferred back-end technology used for the websites. The Wappalyzer library identified 212 back-end technologies from the dataset of 320 web pages. Having in mind the relative abundance of different technologies, according to the data, the distribution of server-side languages is shown in Fig. 1.
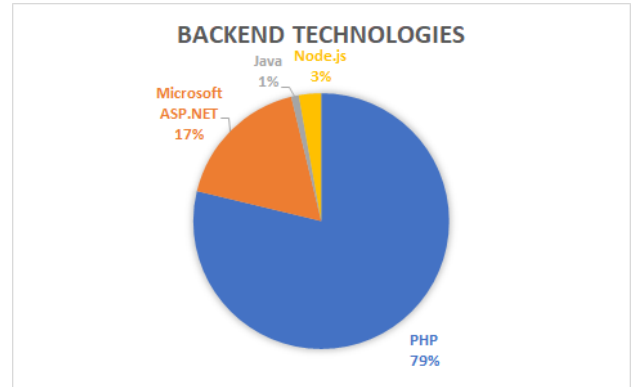

Fig. 1. Distribution of detected programming languages for server-side development.

As can be seen, the leading choice for a server-side programming language is PHP, the second is .NET family of technologies, while Node.js and Java are less used. PHP is the most widely used programming language for back-end development, mainly because of the abundance of free and production-grade software like CMS software, forums, etc. The share of PHP is similar to the various global trends while the share of .NET technologies are significantly larger than the worldwide trends (see [8] for worldwide statistics). Interesting trends may be inferred from the data at this point. For example, besides the 4 most used technologies, there is a notable absence of the less popular technologies, like Ruby and Scala. Also, according to the mapping of the IT industry in Macedonia, published by IT.mk website [9], there is a disparity in the technologies that are used by developers. This can be attributed to the outsource-scope of the Macedonian IT workforce.

In a similar manner, the library was able to correctly identify 205 web servers in the productions. Here, without any surprises, Apache (httpd) is first, while Nginx comes a close second, IIS is trailing third, LiteSpeed is fourth and the Cowboy/Kestrel/OpenRestry/OpenGSE combo comes in fifth, each one used only by one website (Fig.2).

The next question that arises is the number of websites that are using general-purpose software like CMS (Content Management System) and its distribution. Figure 3 shows the dominance of WordPress as the CMS of choice for the web,
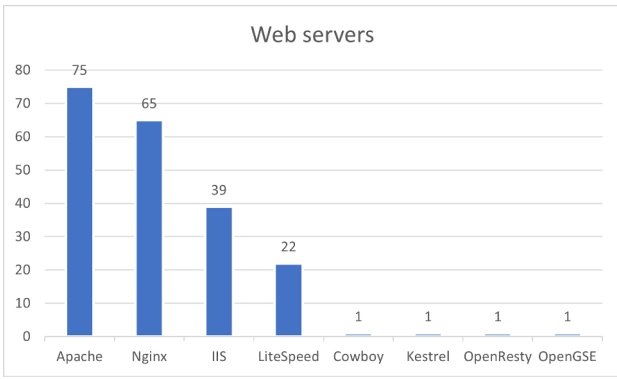
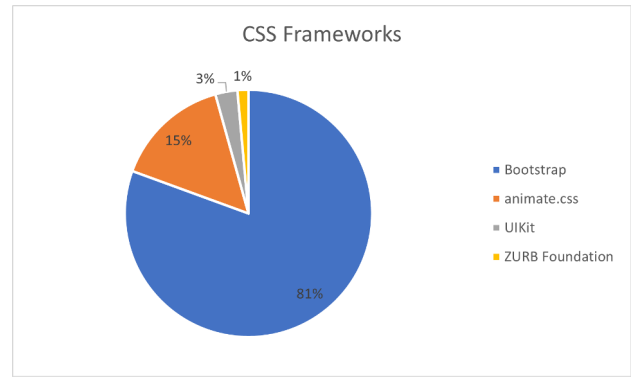Fig. 2. Distribution of web server software.

with 86 percent. It should be noted that only 162 websites were identified as having a well-known CMS solution.
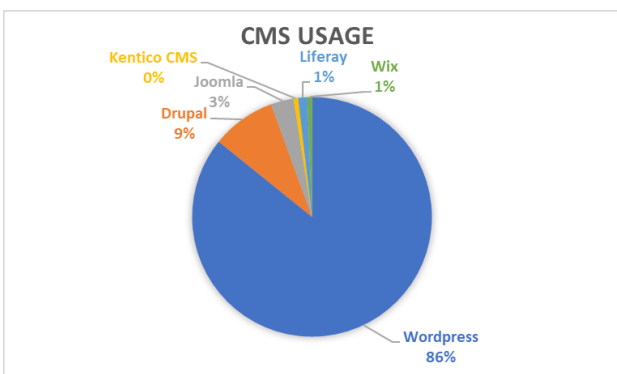

Fig. 3. Proportions of content management systems (CMS)

As further insight into the back-end technologies is limited, attention will be directed towards the front-end technologies. While the front-end can be divided into different aspects, this paper will concentrate on the most common division - visuals and functionality.

The primary visual identity of the page is usually a function of CSS (Cascading Style Sheets), typography and interactions/animations.

With more than 80 percent, Bootstrap is the leading CSS framework for designing the visual identity of websites. In figure 4 is shown that *animate.css* comes second, but it is not a standalone solution for developing visual identity but works along with other CSS frameworks. UIKit and Foundation represent a minuscule part of the CSS frameworks.

When discussing typography, various factors can be analyzed, such as default fonts, number of fonts, font libraries, and more. The subsequent section will focus on presenting the services that are most commonly used to serve web fonts (Fig. 5). Google Fonts and FontAwesome are dominant in its usages. It should be noted that the libraries are not mutually exclusive, meaning that all of these font libraries can be used on a single website.

Functionality-wise, the utilized JavaScript-based technologies that enhance the website functionality can be mapped.

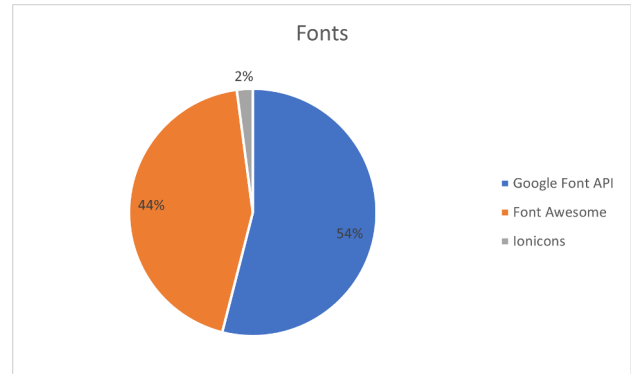Fig. 4. Bootstraps dominates the front-end CSS frameworks.

Fig. 5. Usage of web-font libraries.

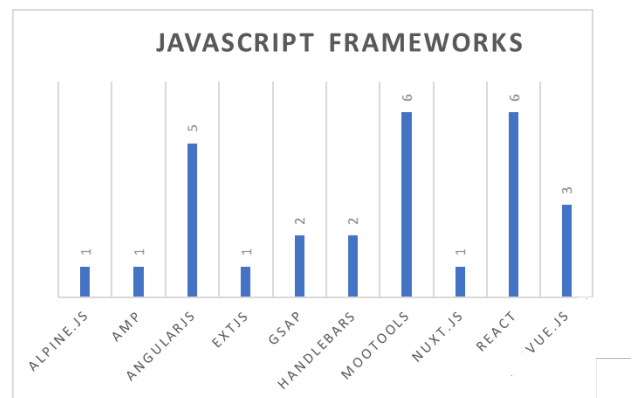The distribution of front-end frameworks is given on Fig. 6.

Fig. 6. Use of front-end frameworks

As can be seen, the relative usage of front-end frameworks is tiny in comparison to the dataset of websites. Only 34 websites have been identified with some form of an front-end framework or around 10 percent of the analyzed websites. Interestingly, *mootools*, a relatively old and hardly updated framework has the same number of users as the most popular, React. Angular is also a popular choice, as well as Vue. Excluding mootools, the distribution of the front-end frameworks

is more-or-less same with the popularity of the technologies worldwide, but the small number of websites that are using it came as a surprise.

When trying to find the most used js library, and having the data about the front-end framework, it came as no surprise that jQuery is the most used JavaScript library, which is visualized in figure 7. Its presence was detected on more than 230 websites, making it the most used JavaScript library in the Macedonian web space. The second place is taken by jQuery Migrate, a library for easing the migration of old jQuery features to a newer jQuery version. The other libraries are specific to some functionality, i.e. *Slick* is the predominant library for displaying carousels.
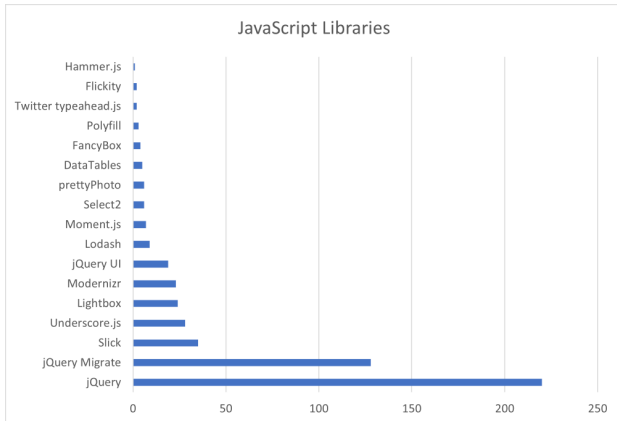

Fig. 7.   Usage of JavaScript Libraries

The last functionality that will be discussed is the preferred usage of Captcha, shown in figure 8. Google's reCaptcha is dominant, while the newer hCaptcha has a single-digit share. This is to be expected, as reCaptcha is an older and better-known service than hCaptcha.

Fig. 8.   Captcha providers for Macedonian websites

V. CONCLUSION AND FURTHER WORK

This paper presents a preliminary investigation of the Macedonian web space, analyzing its trends and technologies while exploring how it is maintained. The widespread use of free technologies like PHP was initially anticipated due to their abundance of free and production software. However, it was unexpected to find that modern front-end frameworks were rare, and those that were implemented were mostly outdated. Additionally, this study revealed the absence of Ruby in back-end technologies and the strong position of .NET.

Further research is necessary to identify potential vulnerabilities associated with outdated libraries and frameworks, as well as to investigate the accessibility features offered by these websites to ensure that they are accessible to all users. The findings highlight the need for ongoing research and development in web technologies to keep up with the ever-evolving web landscape

REFERENCES

[1] N. A. Rakhmawati, S. Harits, D. Hermansyah, and M. A. Furqon, "A SURVEY OF WEB TECHNOLOGIES USED IN INDONESIA LOCAL GOVERNMENTS," *SISFO*, vol. 7, no. 3, pp. 213–222, May 2018, number: 3. [Online]. Available: https://elib.its.ac.id/jurnal/jurnalsisfo/submit/index.php/sisfo/article/view/30

[2] P. A. Nugroho and H. Steven, "Measurement of Unsupported Applications used in Indonesia Popular Websites," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 7, no. 1, p. 18, Apr. 2021. [Online]. Available: http://journal.uad.ac.id/index.php/JITEKI/article/view/18512

[3] S. Kaur, K. Kaur, and P. Kaur, "Analysis of Website Usability Evaluation Methods," Dec. 2018.

[4] D. E. Alvarez, D. B. Correa, and F. I. Arango, "An analysis of XSS, CSRF and SQL injection in colombian software and web site development," in *2016 8th Euro American Conference on Telematics and Information Systems (EATIS)*, Apr. 2016, pp. 1–5.

[5] M. Binder, "Amazon is shutting down one of the oldest staples on the internet," Dec 2021. [Online]. Available: https://mashable.com/article/amazon-is-shutting-down-alexa-internet

[6] Find out the technology stack of any website. create lists of websites and contacts by the technologies they use. [Online]. Available: https://www.wappalyzer.com/

[7] M. Salvati (@byt3bl33d3r), Clay McClure, "python-wappalyzer: Python implementation of the wappalyzer web application detection utility." [Online]. Available: https://github.com/chorsley/python-Wappalyzer

[8] (2023) Usage statistics of server-side programming languages for websites. [Online]. Available: https://w3techs.com/technologies/overview/programming$_{language}$/

[9] M. LLC. (2019) Mapping the it industry in north macedonia 2019. [Online]. Available: https://it.mk/mapiranje-na-it-industrijata-vo-mk-en/