



## Article

# Aesthetic Trends and Semantic Web Adoption of Media Outlets Identified through Automated Archival Data Extraction

Aristeidis Lamprogeorgos, Minas Pergantis \*, Michail Panagopoulos and Andreas Giannakouloupoulos \*

Department of Audio and Visual Arts, Ionian University, 7 Tsirigoti Square, 49100 Corfu, Greece; a18labr@ionio.gr (A.L.); mpanagop@ionio.gr (M.P.)

\* Correspondence: a19perg6@ionio.gr (M.P.); agiannak@ionio.gr (A.G.)

**Abstract:** The last decade has been a time of great progress in the World Wide Web and this progress has manifested in multiple ways, including both the diffusion and expansion of Semantic Web technologies and the advancement of the aesthetics and usability of Web user interfaces. Online media outlets have often been popular Web destinations and so they are expected to be at the forefront of innovation, both in terms of the integration of new technologies and in terms of the evolution of their interfaces. In this study, various Web data extraction techniques were employed to collect current and archival data from news websites that are popular in Greece, in order to monitor and record their progress through time. This collected information, which took the form of a website's source code and an impression of their homepage in different time instances of the last decade, has been used to identify trends concerning Semantic Web integration, DOM structure complexity, number of graphics, color usage, and more. The identified trends were analyzed and discussed with the purpose of gaining a better understanding of the ever-changing presence of the media industry on the Web. The study concluded that the introduction of Semantic Web technologies in online media outlets was rapid and extensive and that website structural and visual complexity presented a steady and significant positive trend, accompanied by increased adherence to color harmony.

**Keywords:** media industry; online media; news websites; Web data extraction; Semantic Web; Web aesthetics; Web archives; World Wide Web



**Citation:** Lamprogeorgos, A.; Pergantis, M.; Panagopoulos, M.; Giannakouloupoulos, A. Aesthetic Trends and Semantic Web Adoption of Media Outlets Identified through Automated Archival Data Extraction. *Future Internet* **2022**, *14*, 204. <https://doi.org/10.3390/fi14070204>

Academic Editor: Michael Sheng

Received: 8 June 2022

Accepted: 29 June 2022

Published: 30 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Journalism was one of the first fields to make the transition from the physical realm to the online digital space, starting with the appearance of the Wall Street Journal in Bulletin Board Systems of the 1980s [1]. As soon as the World Wide Web started becoming popular, newspapers also started being published online, with the Palo Alto Weekly being available on the Web as early as January 1994 [1]. In the beginning, the printed content was being identically reproduced on the Web, but after a short period, some publications started being produced specifically for the Web, thus dramatically changing the way media outlets produced and disseminated their content according to Karlsson and Holt [2]. By the year 1999, more than 20% of American online newspaper content were Web originals, as claimed by Deuze's research the same year [3]. Ever since then, online media outlets have been capitalizing on the Web's power to provide journalistic content with traits that only it can offer, namely interactivity, immediacy, hypertextuality, and multimodality [2].

At the turn of the millennium, Tim Berners-Lee proposed the Semantic Web, an expansion of the World Wide Web that included content that can be retrieved and comprehended by machines, introducing the idea of a machine-readable Web [4]. In the field of Journalism, as Fernandez et al. point out, in order to cover the customers' needs for information freshness and relevance, the use of metadata became prevalent [5]. Moreover, the use of additional Semantic Web technologies as proposed by Fernandez et al. was set to increase both productivity and media outlet revenues [5]. Heravi and McGinnis proposed the use

of Semantic Web technologies, in tandem with Social Media technologies, to produce a new Social Semantic Journalism framework that combined technologies that could collaborate with each other in order to identify newsworthy user-generated journalistic content [6].

However, the evolution of the Web is not limited to content diffusion and machine-readability fields but also applies to the realms of aesthetics and usability. As Wu and Han point out, both aesthetics and usability display a strong relationship with the satisfaction of potential users [7]. King et al. [8] make the claim that a significant relationship exists between the visual complexity of a website and its influence on user first impressions. This is especially important with regard to media outlets since King's research specifically links increased visual complexity with the user's perception of informativeness and engagement cues [8]. This perceived informativeness is an important quality when associated with a news website. Besides complexity, usability and compatibility with multiple devices have also evolved through the progression of website layout techniques over the course of time as studied by Stoeva [9]. The way information is presented on a Web page is under constant change.

In addition to complexity and layout, color also plays an important role in influencing user impressions. On many occasions, researchers have established that the colors used on a website can elicit emotional reactions and feelings that can lead to outcomes concerning a website's perceived trustworthiness and appeal or even a visitor's overall satisfaction [7,10–12]. Talei proposes that these emotional responses are a result of human natural reactions to colors as encountered in natural life [12]. In addition to colors as individual factors eliciting an emotional reaction from users, White proposes that color schemes can also have a similar effect and proceeds to study the case of schemes using complementary colors [13] leading to conclusions about how specific complementary colors lead to increase in user pleasure.

In order to monitor how websites of media outlets evolve alongside the evolution of Web technologies and aesthetics, taking a look at contemporary websites only is not enough. Instead, what is needed is a comprehensive overview of each website's journey throughout the past decades. Brügger coined the term "website history" as a combination between media history and Internet history, where the individual website is considered the object of historical analysis instead of the medium [14]. The website then, playing the part of a historical document, is to be archived and preserved, and subsequently delivered as historical material [14]. This type of historical material is the means through which the aesthetic trends and Semantic Web adaptation of media outlets may be identified through means of archival data extraction.

The study presented in this article attempts to answer the following research questions:

RQ1. How has the integration of Semantic Web technologies (SWT) progressed in the last decades? When and to what extent were various technologies implemented?

RQ2. What are the trends in website aesthetics that can be identified concerning the complexity of Web pages, the usage of graphics, and the usage of fluid or responsive designs?

RQ3. What basic colors and coloring schemes are prevalent in website homepages? Did they change over the years and are there consistent trends that can be inferred by such changes?

In order to investigate these questions, large amounts of quantitative data were collected from actual public media outlets on the World Wide Web, based on their popularity in Greece. The past versions of these websites were retrieved through the use of a Web service offering archival information on websites. With that data in hand, a comprehensive understanding of the landscape of SWT adoption and general aesthetic trends can be attained. The method of collecting and analyzing that information will be presented in the following section.

## 2. Methodology

The research presented was conducted in four stages:

Stage 1: Media outlet websites were identified and selected based on their popularity in Greece.

Stage 2: Current and archival information from these websites was collected through the use of a website archive service. This information included the HyperText Markup Language (HTML) code of a website’s homepage as well as a screenshot of that homepage.

Stage 3: Using a Web data extraction algorithm, information regarding the usage of SWTs, website complexity, graphic usage, and website repressiveness or fluidity was recorded.

Stage 4: Using an image analysis algorithm, information regarding the colors used was extracted from the websites’ screenshots.

The methods and decision process behind each stage will be further detailed in this section. The quantitative data collected will be further presented in the results section.

### 2.1. Identifying Websites for Information Extraction

In order to reach safe conclusions regarding the evolution of media outlet websites through time, a large number of websites must be used, as well as multiple instances of each such website over the course of time. A large data set can lead to reliable results and create an impression that accurately represents reality. For that purpose, the archival Web service that was selected as the main provider of data concerning these websites was the Wayback Machine Internet Archive. As seen in the work of Gomes et al. [15], most Web archiving initiatives are national or regional. Out of the few international ones, the Internet Archive is both the largest and the oldest, dating back to 1996. It boasts over 625 million Web pages [16] which it provides to interested parties through its Wayback Machine. Using the Wayback Machine was considered the best way to collect a variety of instances for each studied website, which spanned over a representative period of time.

Another consideration, besides the number of instances, was which specific websites were to be targeted. A reliable metric of a media outlet’s impact and visibility is its popularity based on digital traffic. Additionally, this popularity can ensure the existence of multiple instances of archived website data in Web archives. Based on that, a sample of the 1000 most popular websites was obtained from the SimilarWeb digital intelligence provider in the category of “News & Media Publishers” in Greece. SimilarWeb is a private company aiming to provide a comprehensive and detailed view of the digital world [17]. Information about a website’s online market share, its global rank, and more were collected manually in the form of text files and using an algorithm scripted with PHP, this information was parsed and imported into a relational database powered by the MariaDB database management engine. This process is visually presented in Figure 1.

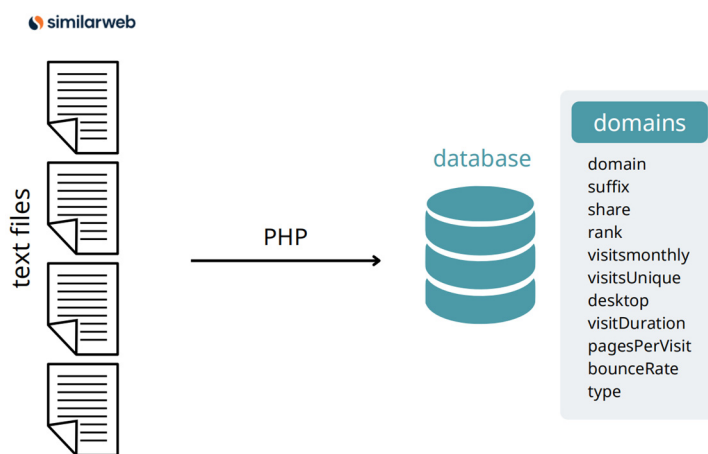


Figure 1. Visual representation of the process of website information collection.

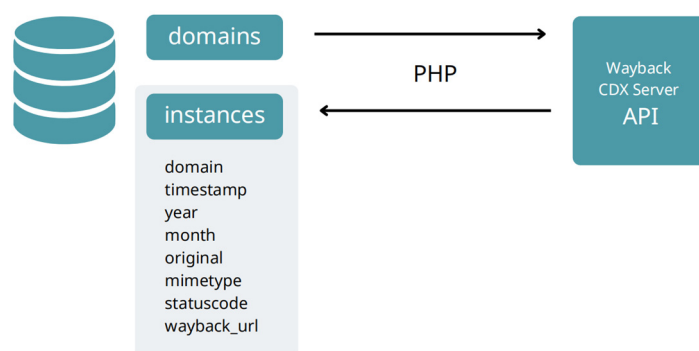
Both international websites with a popular presence in Greece and popular Greek media outlets were included in the final list of websites to be investigated. Overall, the websites presented a varied mix including popular international online media outlets

(e.g., Yahoo, MSN, BBC, the NYTimes, etc.), popular Greek online media outlets such as (e.g., protothema.gr, iefimerida.gr, newsbomb.gr, etc.), a series of local news outlets with a popular online presence (e.g., typosthes.gr, thebest.gr, larissanet.gr, etc.), and more.

## 2.2. Collecting HTML Data and Screenshots of Each Relevant Website

Having established a good dataset of relevant websites, the next stage of this research was to collect HTML data and screenshots for each website for various different instances over the past few decades. An algorithm was developed in the PHP scripting language that inquired the Internet Archive's Wayback Machine for each of the websites collected in the previous stage, in order to obtain available instances for that specific website.

These inquiries were performed using the Wayback CDX (ChemDraw Exchange format) server Application Programming Interface (API). The CDX API is a tool that allows advanced queries that can be used to filter entries with high density instancing, in order to obtain instances for specific intervals. By using the API's ability to skip results in which a specific field is repeated, instance recovery was accomplished faster and more efficiently. For each instance of a website that is discovered in the Internet Archive's database, the API provides information on the domain name, the exact timestamp of the snapshot, the snapshot's year and month, the original Uniform Resource Locator (URL), the mime type of the data provided by the service and the current URL of the archived website on the Wayback Machine. This process of collecting instances is visually presented in Figure 2.



**Figure 2.** Visual representation of the process of instance information gathering.

Koehler, in their research, discovered that the half-life of a page is approximately 2 years [18]. Especially when it comes to structural or large-scale changes such as the ones that are being investigated in this research, it makes sense that they do not happen too often. With that in mind, for the purposes of this study, it was decided that one website instance per year was more than enough to record any significant changes. In order to accomplish this sampling, the timestamp field that was returned by the API was utilized. This field has 14 digits corresponding to the year, month, day, hour, minutes, and seconds that the instance was created. By instructing the API to exclude results that had the same first four digits in this field, the system returns exactly one snapshot per year as intended (if available). Out of a total of 1000 websites identified in stage one, 905 were discovered in the Internet Archive's databases and a grand total of 10,084 instances were discovered.

In order to acquire the HTML source code for each instance, an algorithm was developed in the PHP scripting language. This algorithm made use of the Wayback URL field that was collected during the instance information-gathering process to access the archival version of the website on the Wayback Machine. After accessing the instance, the algorithm proceeded to extract the source code and store it in an HTML file. The files were stored in a separate folder for each domain and their filenames represented the year and month of the instance. Before storing the source code into the HTML file, the application used string manipulation PHP functions to remove any part that belonged to the Wayback Machine's Web interface, in order to ensure that the end result was exclusively the original website's source code. This process is visually presented in Figure 3.

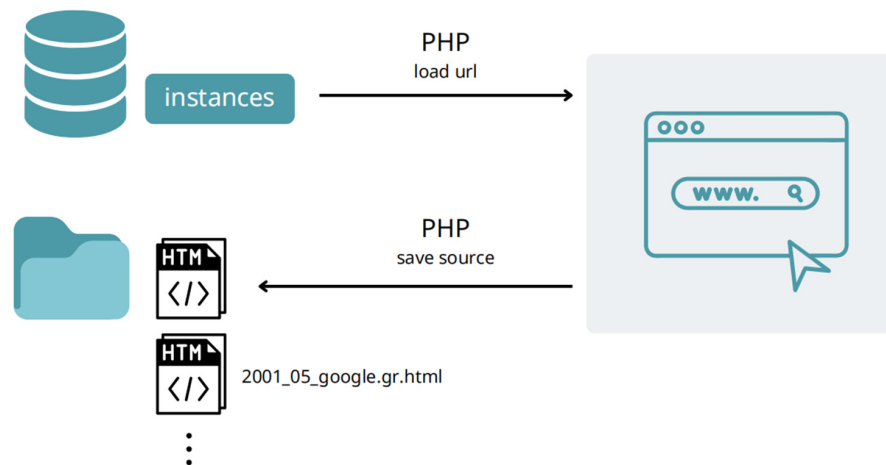


Figure 3. Visual representation of the HTML code collecting process.

The second important piece of information collected in this stage of our research besides the HTML source code is a screenshot of each website instance’s homepage. The collected screenshots will be used to infer the color pallets of each instance and derive information from there. The plug-in UI.Vision RPA for the Chrome browser was used to acquire these screenshots. This plugin is a tool that allows the automation of various browser operations. The instructions for the automated process are provided to the plugin using JSON syntax. This enabled us to generate a vast series of instructions using an algorithm in the PHP scripting language. These instructions guide the plugin to open a website, pause for the time required for the website to load, capture a screenshot of the website, and then proceed to store the captured screenshot in a PNG image file with a filename indicating the year, month, and domain of the instance. This process is visually presented in Figure 4.

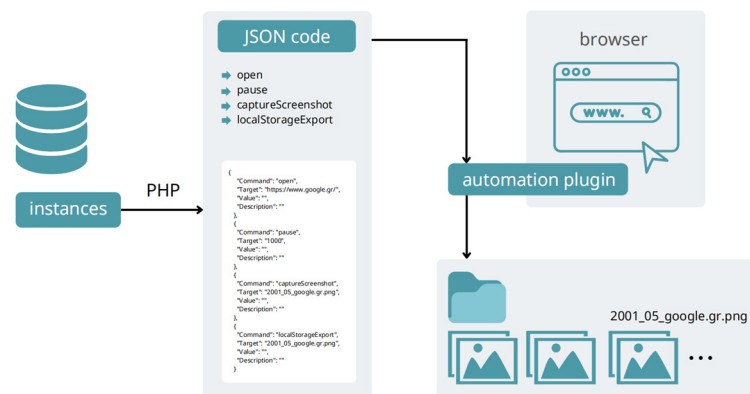
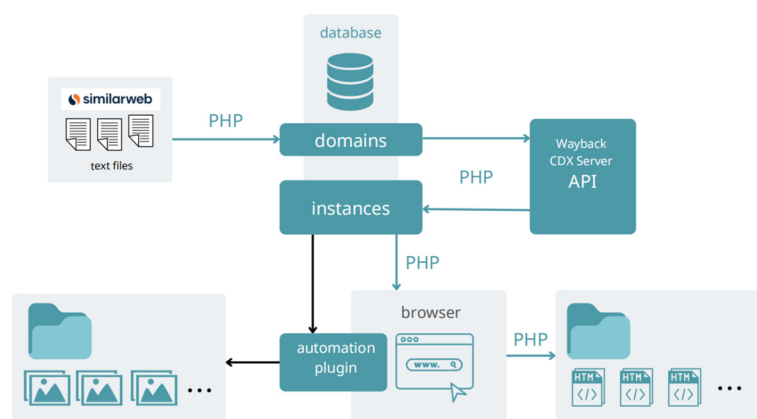


Figure 4. Visual representation of the screenshot collecting process.

This detailed overall process of gathering archival data related to website aesthetics can be extended for use in other fields and with different objectives that can be accomplished through knowledge of the HTML source code and a screenshot of a website instance and was presented in greater detail by Lamprogeorgos et al. in 2022 [19]. The complete process is visually presented in Figure 5. It should be noted that the process of collecting screenshots is much more resource and time intensive than the process of collecting HTML documents and for this reason, the analysis of screenshots was based on a random sample of 5402 website instance screenshots out of the 10084 total website instances. The screenshot sample was considered still large enough to lead to safe conclusions.





**Figure 5.** Visual representation of the complete instance data collecting process.

### 2.3. Collecting SWT and Aesthetics Data from the HTML Source Code

With the HTML files containing the source code of each website instance collected, the next step included the extraction of data from these files. The process of collecting this information was accomplished with the use of an algorithm developed in the PHP scripting language. This algorithm converted each HTML file into an entire HTML document through the use of PHP's DOMDocument Class. It then proceeded to collect information based on the various HTML elements and their attributes. This information was recorded into variables that can be divided into three categories: variables concerning Semantic Web technology adoption, variables concerning the homepage's complexity, and variables concerning the user interface's layout.

#### 2.3.1. Semantic Web Technologies Adoption Variables

With the coming of HTML5 in 2008, a series of new structural elements were added [20] with the intention of not only providing structural insight, as normal HTML elements do but also contextual insight on what the content inside these elements represents. Fulanovic et al. indicate that usage of these elements is mainly intended for browsers and accessibility devices, and that it is up to the content creators to select the proper element to convey the contents of each part of their website [21]. These the elements are `<article>`, `<aside>`, `<details>`, `<figcaption>`, `<figure>`, `<footer>`, `<header>`, `<main>`, `<mark>`, `<nav>`, `<section>`, `<summary>`, and `<time>`. The data extraction algorithm traverses the Document Object Model (DOM) of each website and identifies the use of any of these elements and records it into the variable `html_var`.

The second variable concerning SWT adoption was `og` and it recorded whether a website made use of the Open Graph protocol to present itself in the form of a rich object. The protocol's intention is to make it possible for websites to be presented in a social graph and this is accomplished through a method compatible with W3Cs Resource Description Framework in attributes (RDFa) recommendation [22].

Another RDFa compatible system specifically designed for Twitter is called "Twitter Cards" [23] and whether it existed in a website instance was recorded in the `twitter` variable. Both the Open Graph and the Twitter Cards graphs create meta tag attributes that include information containing the Web page including a title, a short description, and a related image. Essentially, both Open Graph and Twitter Cards comprise Semantic Web applications that stem from the realm of Social Media as Infante-Moro et al. explain [24] and this connection they have with Social Media has influenced their popularity and their importance to websites' Semantic Web integration.

Although technically on the fence between Web 2.0 and Web 3.0, RDF Site Summary (RSS) feeds present one of the earliest attempts at Web syndication [25] and have hence been a long-time component of presenting Web pages and their content in a machine-readable manner. The variable `rss` records the existence of such feeds in a website instance.

Finally, the last SWT-related variable is *sch*, which records the existence of schema.org data structures in the website instance. The data structure schemas of the schema.org community, which is supported by various big names in Web technologies such as Microsoft, Google, Yahoo, and Yandex, aim to make it easier for developers to integrate sections of machine-readable information in their creations [26]. Their usage provides the flexibility of choosing between three formats: RDFa like Open Graph and Twitter Cards, Microdata, and JSON-LD.

Table 1 presents all SWT-related variables with a short description.

**Table 1.** Variables related to Semantic Web technologies.

Variable Name	Description
<i>html_var</i>	Records semantic HTML5 elements
<i>og</i>	Records the existence of an Open Graph RDFa graph
<i>twitter</i>	Records the existence of a Twitter Cards RDFa graph
<i>rss</i>	Records the existence of an RSS Feed
<i>sch</i>	Records the existence of a schema.org data structure

### 2.3.2. Aesthetics and Interface Variables

Visual complexity is a factor that plays an important role in the aesthetics of a website as discussed by Harper et al. [27], King et al. [8], and Chassy et al. [28]. Harper et al., in their work, supported that complexity as perceived by the users is influenced by structural complexity and presented a paradigm that related the complexity of an HTML document's DOM with how users subjectively judged complexity [27]. In a similar manner, the present study collected information regarding specific DOM elements, including both structural elements and graphical elements, in order to draw conclusions regarding the aesthetics of a website instance and how they evolved through time with regard to visual and structural complexity. In Figure 6, a screenshot of the homepage of popular European media outlet [euronews.com](http://euronews.com) (accessed on 1 January 2018), which displays a high amount of visual and structural complexity, is presented as an example.

In the *div\_tags* variable the number of `<div>` elements was recorded while all hyperlinks were identified through the use of anchor elements `<a>` and recorded in the *a\_tags* variable. Similarly, the various graphical components were measured using the *img\_tags* variable to collect `<img>` elements, the *svg\_tags* variable to collect scalable vector graphics elements (`<svg>`), the *map\_tags* variable to collect image map elements (`<map>`), the *figure\_tags* variable to collect figure semantic element (`<figure>`), the *picture\_tags* variable to collect the art and responsive design oriented picture element (`<picture>`), and finally the *video\_tags* variable to collect `<video>` elements.

The `<img>` tag is used to embed an image file in an HTML page. The image file can be of any Web-supported filetype such as compressed JPG files, animated GIF files, transparent PNG files, and even SVG files. An SVG element (`<svg>`) is a graphic saved in a two-dimensional vector graphic format that stores information that describes an image in text format based on XML. An image map consists of an image with clickable areas, where the user can click on the image and open the provided destination. The `<map>` tag can consist of more than one `<area>` element, which defines the coordinates and type of the area and any part of the image can be linked to other documents, without dividing the image. The `<figure>` tag is used to mark up a photo in the document on a Web page. Although the `<img>` tag is already available in HTML to display the pictures on Web pages, the `<figure>` tag is used to handle the group of diagrams, photos, code listing, etc. with some embedded content. The most common use of the `<picture>` element will be in responsive designs where instead of having one image that is scaled up or down based on the viewport width, multiple images can be designed to more nicely fill the browser viewport.

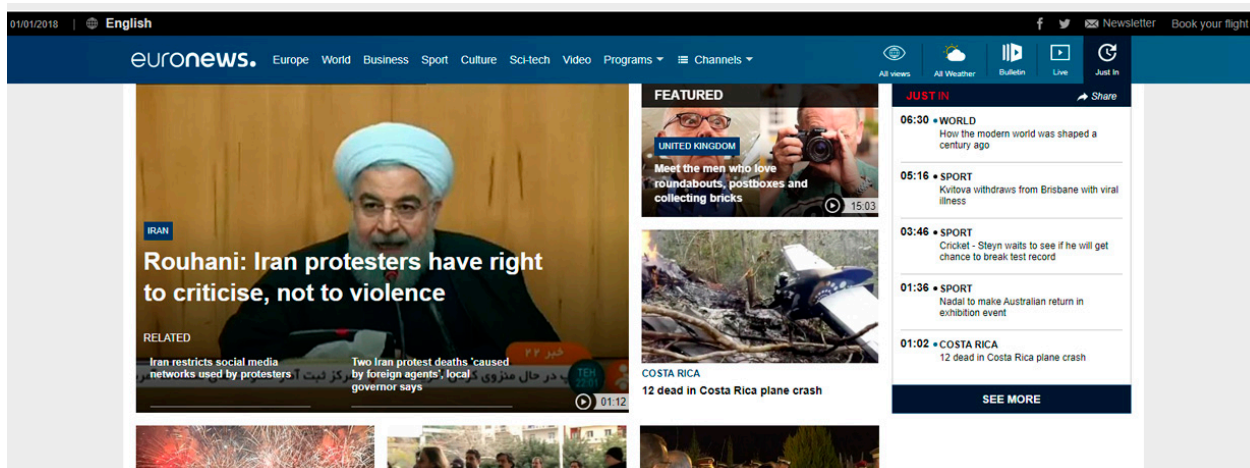


Figure 6. Example of the homepage of euronews.com displaying increased structural and visual complexity.

Besides visual complexity, modern website aesthetics and their interfaces are heavily influenced by the need to be presentable and easily usable on many different devices, operating at various different screen resolutions and aspect ratios. This has been achieved through the fluidity offered by using table elements to contain a website’s structure and through the use of responsive design practices and frameworks. In order to study the trends in this area over time for each website instance, the number of table elements (<table>) was recorded in the *table\_tags* variable. Additionally, the viewport meta element was investigated for each website instance as an indicator that the website is undertaking an effort towards supporting multiple screen resolutions and the results were recorded in the *mobile\_scale* variable. Finally, two very popular responsive design frameworks were investigated. These were Bootstrap, an open source CSS framework developed by the Bootstrap Team and operating under the Massachusetts Institute of Technology (MIT) license [29], and Foundation, a similar CSS framework also operating under the MIT license developed by ZURB [30]. In order to identify the frameworks, the algorithm tried to detect div elements with the grid “row” class and then proceeded to investigate for grid column elements through the various “col-” classes for Bootstrap and the “columns” and “large-” or “small-” classes for Foundation. Whenever the use of these frameworks was discovered, it was recorded in the *bootstrap* and *foundation* variables respectively.

Table 2 presents all visual complexity and layout structure-related variables with a short description.

Table 2. Variables related to visual complexity and layout structure.

Variable Name	Description
<i>div_tags</i>	Records the number of <div> elements
<i>a_tags</i>	Records the number of <a> elements
<i>img_tags</i>	Records the number of <img> elements
<i>svg_tags</i>	Records the number of <svg> elements
<i>map_tags</i>	Records the number of <map> elements
<i>figure_tags</i>	Records the number of <figure> elements
<i>picture_tags</i>	Records the number of <picture> elements
<i>video_tags</i>	Records the number of <video> elements
<i>table_tags</i>	Records the number of <table> elements
<i>mobile_scale</i>	Records the number of <div> elements
<i>bootstrap</i>	Records the existence of elements with classes used by the bootstrap framework
<i>foundation</i>	Records the existence of elements with classes used by the Foundation framework



### 2.4. Collecting Color Data from the Homepage Screenshot

Having amassed a large amount of website instance screenshots, we proceeded to use them in order to gain a better understanding of how news websites evolved through the last decades, in terms of empty space use and colors. Empty space (or white space, or negative space) is the unused space around the content and elements on a website, which designers used to balance the design of the website, organize the content and elements, and improve the visual experience for the user. Figure 7 presents an example of empty space from a homepage screenshot from the popular American media outlet [nytimes.com](http://nytimes.com) (accessed 1 January 2014), where all the empty space has been marked with the use of the color orange.



Figure 7. Example of a homepage screenshot from [nytimes.com](http://nytimes.com) with the empty space turned orange.

Figure 8 displays an example of the evolution of the homepage of the international media outlet [hellomagazine.com](http://hellomagazine.com) throughout the last two decades. This collection of homepage screenshots exemplifies the visible evolution of structural and graphical complexity, as well as color and empty space usage, which comprise the metrics collected by our algorithms from each website instance, as detailed in Section 2.3.2 and in the current section.

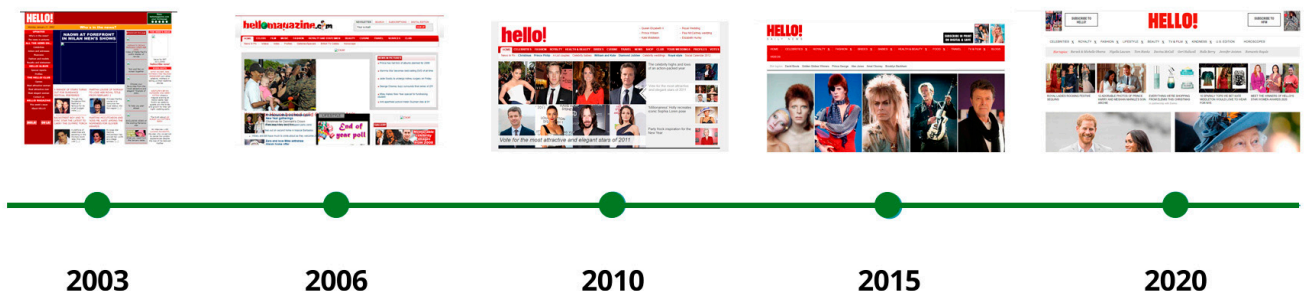


Figure 8. The evolution of the homepage of [hellomagazine.com](http://hellomagazine.com) throughout the last two decades.

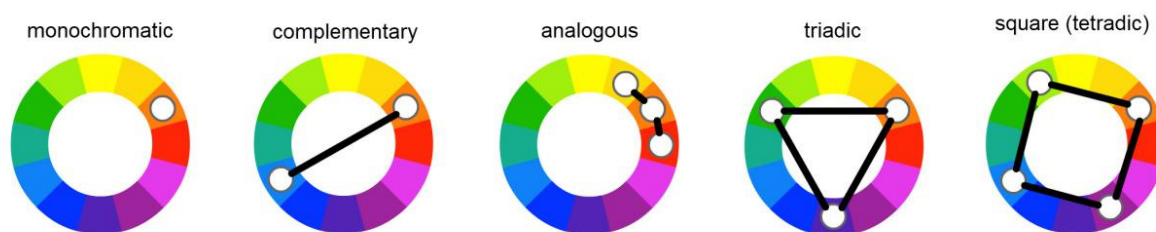
An algorithm was created that used the PHP scripting language and its native image handling capabilities to discover information regarding the use of color as presented by the screenshots. At first, the algorithm used image scaling and the *imagecolorat* function to identify and extract colors from a screenshot into the hexadecimal color code used by HTML5 and CSS3. Our work was based on the ImageSampler class developed by the Art of the Web [31]. All colors that took less than 3% of space on the screenshot are excluded from further analysis. In order to better study the remaining extracted colors, they were grouped based on their proximity to a primary, secondary, or tertiary color of the red yellow blue (RYB) color model.

As established by Gage in his work in the 1990s [32] the RYB color model incorporates subtractive color mixing and is one of the most popular color models, especially in design. By extension, it has become very useful in digital art and, of course, Web design since it can be used to identify colors that go well together. A major reason it was decided to convert the red green blue (RGB) based HTML hexadecimal colors to the RYB ones was to better study design schemes based on color relationships, as will be detailed below. The three primary colors of the RYB color wheel are red, yellow, and blue. Each combination of the three creates secondary colors which are orange, green, and purple. The tertiary colors are established through the combination of primary and secondary colors and they are red-orange, yellow-orange, yellow-green, blue-green, blue-purple, and red-purple. Additionally, black is achieved by combining all three primary colors and white through the lack of them.

The algorithm in this research used saturation to determine if a color is white: any color with less than 16% saturation was considered white. In a similar manner, brightness was used to identify black: any color with less than 16% brightness was considered black. Considering websites as a medium are presented across many different types of screens of various technologies, colors that are this close to black or white will most definitely be perceived as such by the average user. Additionally, the most used color on each website instance was considered to be the empty space color, meaning the color upon which all visual elements of the page appear.

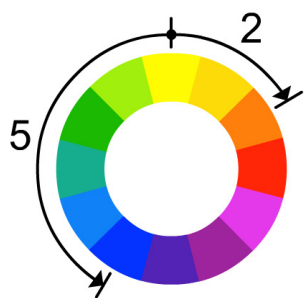
In order to identify whether a color scheme (or color combination) is used in each website instance that uses colors besides black and white, an additional algorithm was developed in the PHP scripting language. This algorithm was designed to identify five major methods of color combination based on the RYB color wheel as presented in Figure 9:

- Monochromatic shades, tones, and tints of one base color
- Complementary colors that are on opposite sides of the color wheel
- Analogous colors that are side by side on the color wheel
- Triadic, three colors that are evenly spaced on the color wheel
- Tetradic, four colors that are evenly spaced on the color wheel



**Figure 9.** Major schemes of color combination based on the RYB color wheel.

The algorithm measured the minimum and maximum distance between the colors on the color wheel. Based on the number of colors and these two distances, conclusions can be drawn regarding the use of a harmonic color combination as presented in Figure 10.



**Figure 10.** Example of measuring distances on the RYB color wheel.

If the number of colors used is one, then the color scheme used is monochromatic. If the number of colors used is two, and if the maximum distance is lower than two, the analogous scheme is used, but if the maximum distance is greater than five, the complementary scheme is used. Similar conclusions can be drawn from the usage of three or four colors. If three colors are used and the minimum distance is greater than three, then the triadic color scheme is used. Similarly, if four colors are used and the minimum distance is greater than two, then the tetradic color scheme is implemented. The algorithm rejects any other situation and classifies it as a non-harmonic color combination.

Having obtained all relevant information through the steps described above, we proceeded to study the following:

- How many colors appear in the website instances on average by year besides black and white?
- How much each of the basic 14 colors of the RYB model is used in the website instances on average by year?
- How popular was the use of white, black, or colored empty space through the years?
- How popular were the different types of harmonic color combination schemes through the years?

The answers to these questions, alongside all other information collected throughout the stages of this research as presented in this section, are available in the results section below.

### 3. Results

The figures presented in this section focus on the various variables measured during the methodology section and how they shifted and changed in the last two decades from 2002 to 2022. Although some data since 1996 is available, it was deemed too small a sample to accurately present the information of that earlier era.

#### 3.1. The Use of Semantic Web Technologies

Using the SWT adoption variable presented in Table 1 the usage of each technology was calculated as a percentage of the total website instances analyzed for each year. A graphic representation of the evolution of SWT adoption is presented in Figure 11. As we can see, the use of RSS feeds was the first introduction of machine-readability-related technologies in popular news media outlets circa 2004. Its use rose steadily in popularity until the early 2010s when the Open Graph technologies started rising to popularity alongside the use of HTML semantic elements. RSS usage stagnated around the same period forming a plateau in its curve. On the other hand, Open Graph and HTML semantic elements continued to rise in popularity until the late 2010s when they seem to have also plateaued but at significantly higher usage percentages. Twitter Cards and schema.org data structures start appearing en masse after 2014 and although their rate of diffusion is lowering they still display an upward trend.

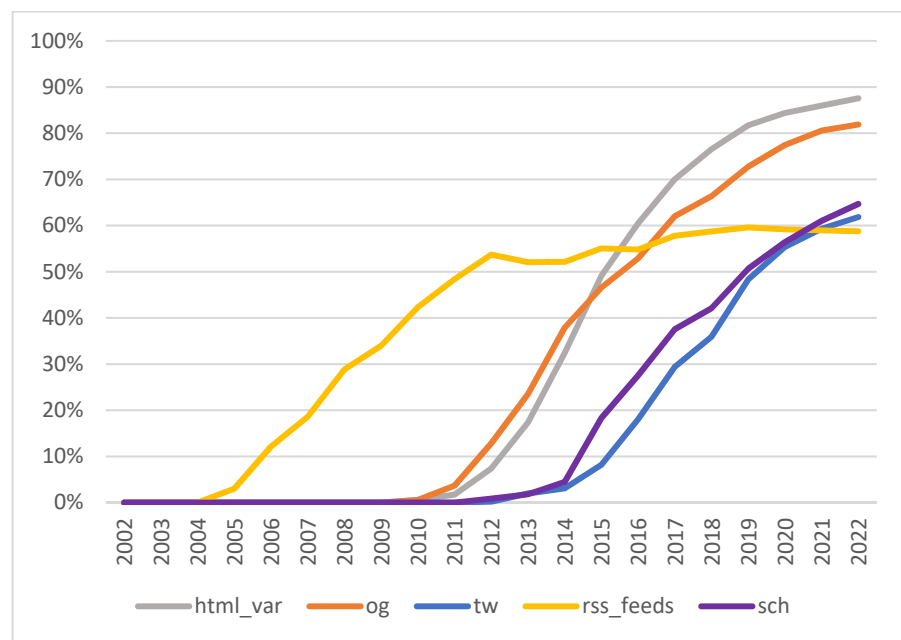


Figure 11. Percentage of websites using various Semantic Web technologies by year.

### 3.2. Website Complexity in Terms of Structural and Graphical Elements

Using the *div\_tags* and *a\_tags* variables as presented in Table 2, an overall impression of the evolution of structural complexity can be achieved. Figure 12 presents two curves, one for each variable, indicating what the average value for each variable was each year. The *div* element is more steep, beginning at a lower starting point in 2002 and ending above. Sometime after 2010, the average number of hyperlinks on news websites’ front pages fell below the average number of *div* elements.

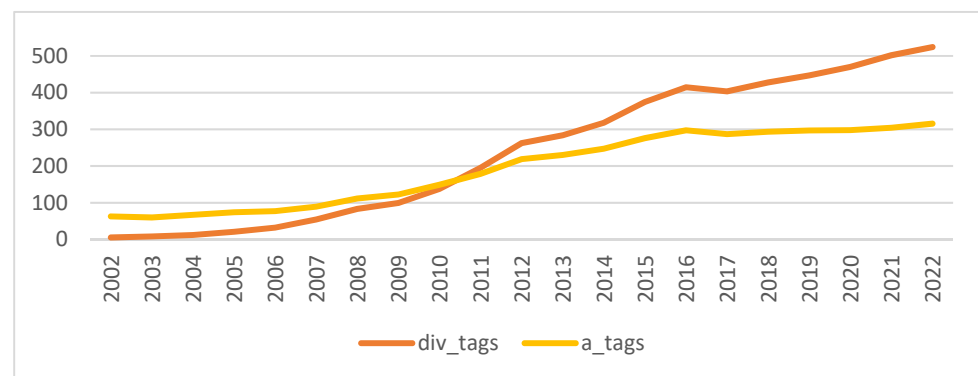


Figure 12. Website complexity as inferred through the average number of hyperlinks and div elements.

The total number of graphical elements identified every year as measured by the *img\_tags*, *svg\_tags*, *map\_tags*, *figure\_tags*, *picture\_tags*, and *video\_tags* of Table 2, were normalized by mapping them between the values of 0% and 100% (100% being the maximum detected number of elements) in order to make them independent of the number of website instances that were investigated, which were different for each year. Figure 13 presents the resulting curves that can be used to infer existing tendencies. Image elements in the form of an *<img>* tag display a slowly increasing trend, on par with the overall complexity increase depicted in Figure 12. From the early 2010s, a sharp rise appears in the figure tags which coincides with the rise of semantic elements usage as presented in Figure 11. The picture element also presents a similarly positive trend but lags 3–4 years behind since it is

more closely related to responsiveness, which will be studied further below. Image maps that were somewhat popular during the 2000s have been in fast decline and are all but extinct in modern websites. On the other hand, scalable vector graphics entered the field in the mid-2010s and their presence has been rising steeply ever since. Finally, the use of the video tag first appears after 2015 and has a peculiar double peak curve with its peaks in 2017 and 2022.

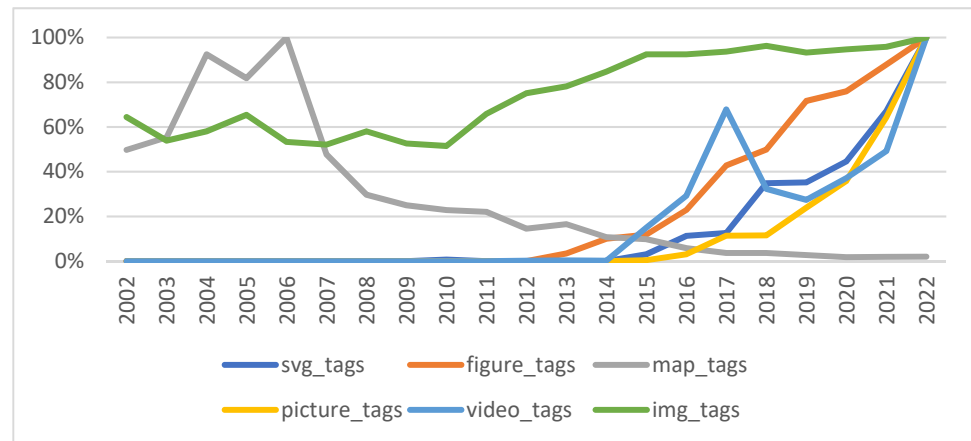


Figure 13. Graph presenting the number of graphical elements (normalized).

### 3.3. Fluidity and Responsiveness in Website Layouts

Fluid design as indicated by the *table\_tags* variable measuring the existence of table elements, mobile device screen support as indicated by the *mobile\_scale* variable, and responsive design as identified in the *bootstrap* and *foundation* variables are depicted in Figure 14. As seen clearly, fluid design after being very popular before 2008, has been in a steady decline since then, while, on the other hand, mobile support and responsive design have been rising rapidly in the 2010s.

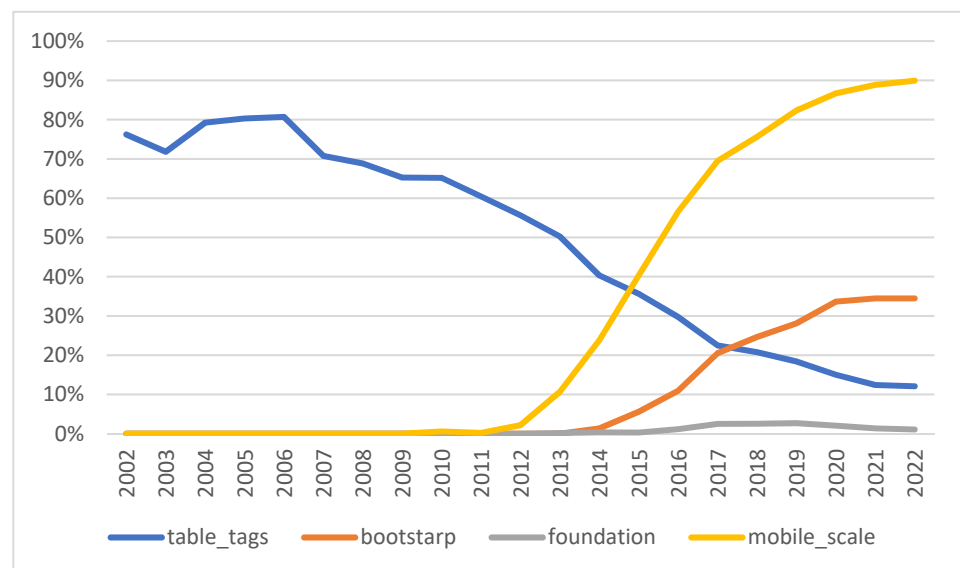


Figure 14. Graph presenting the usage of fluid or responsive design techniques.

### 3.4. Color and Empty Space Analysis Based on Website Homepage Screenshots

The number of different basic colors of the RYB color model that were used in each website instance provides a glimpse of the evolution of color-oriented aesthetical complexity. Figure 15 presents an area chart that depicts what percentage of each year's instances used



multiple colors and to what extent. Zero indicates no colors used besides black or white or other shades of the two, while each other number indicates how many basic colors were used besides black and white. It is made apparent that the use of fewer basic colors is more prevalent, although there is a trend as time goes by for the number of colors used to increase.

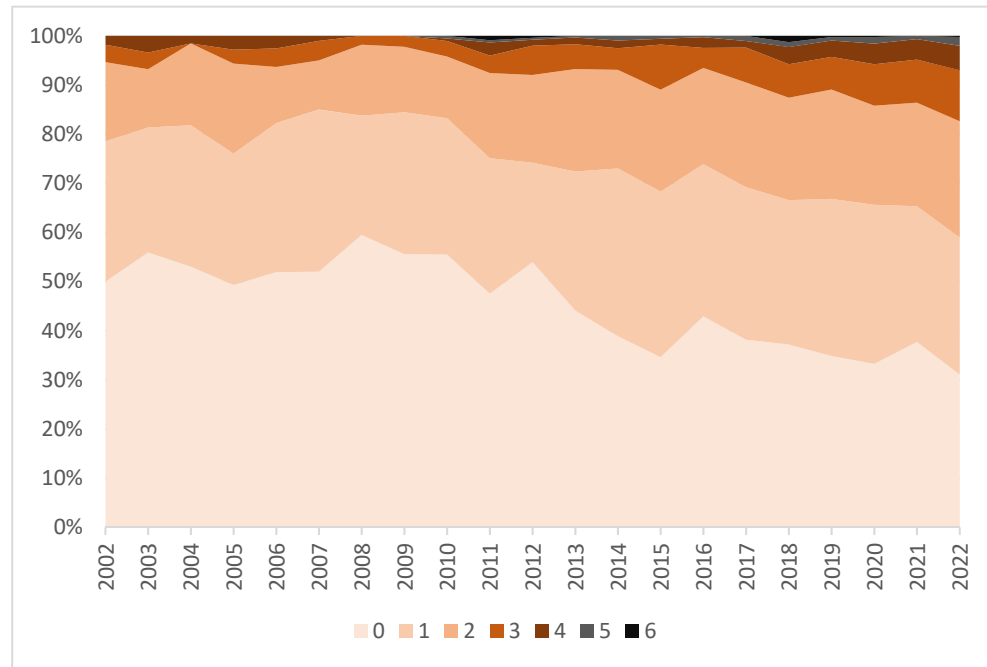


Figure 15. Number of basic RYB colors used besides black and white by year.

In order to form a more in-depth idea of which colors are used most, a graph showing the color usage of all 14 different colors is presented in Figure 16. As expected white is the most popular color, although displaying a somewhat negative trend. On the other side, black seems to display a positive trend, while other colors are much less used.

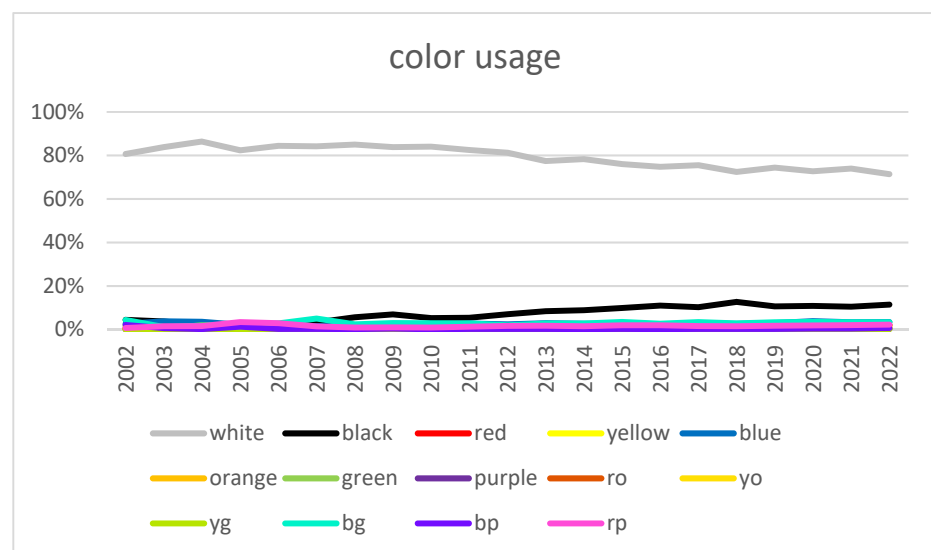
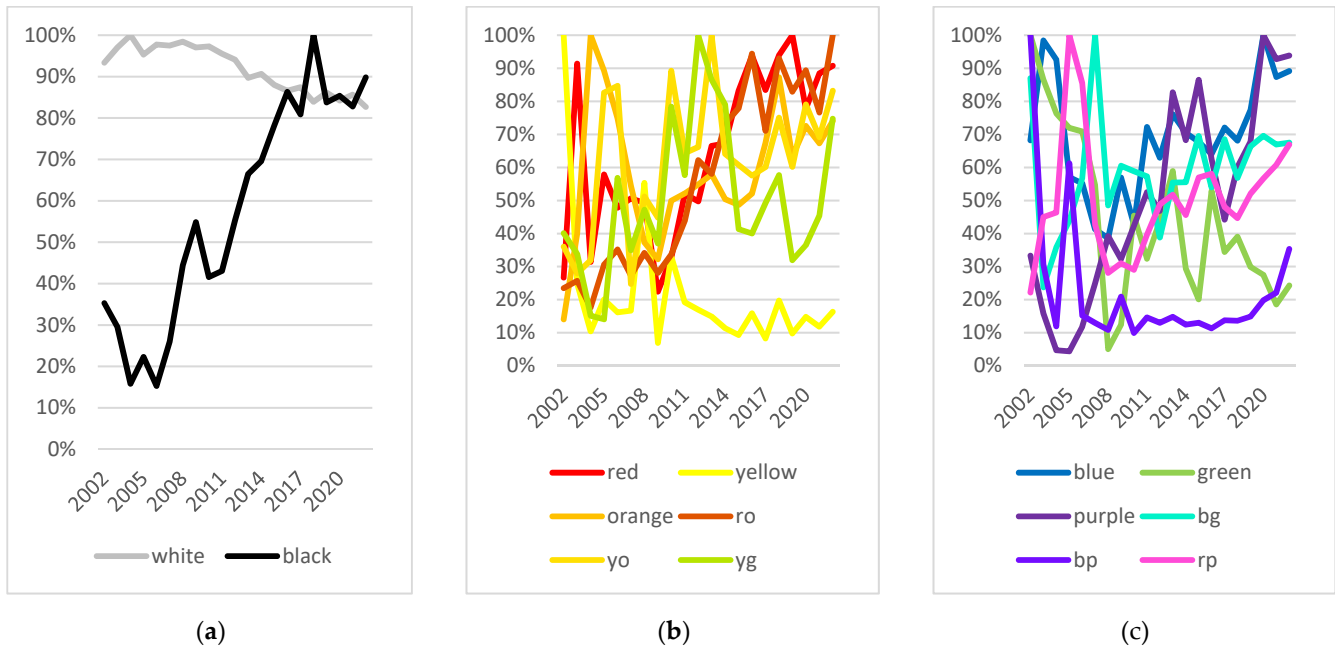


Figure 16. Basic RYB color usage by year.

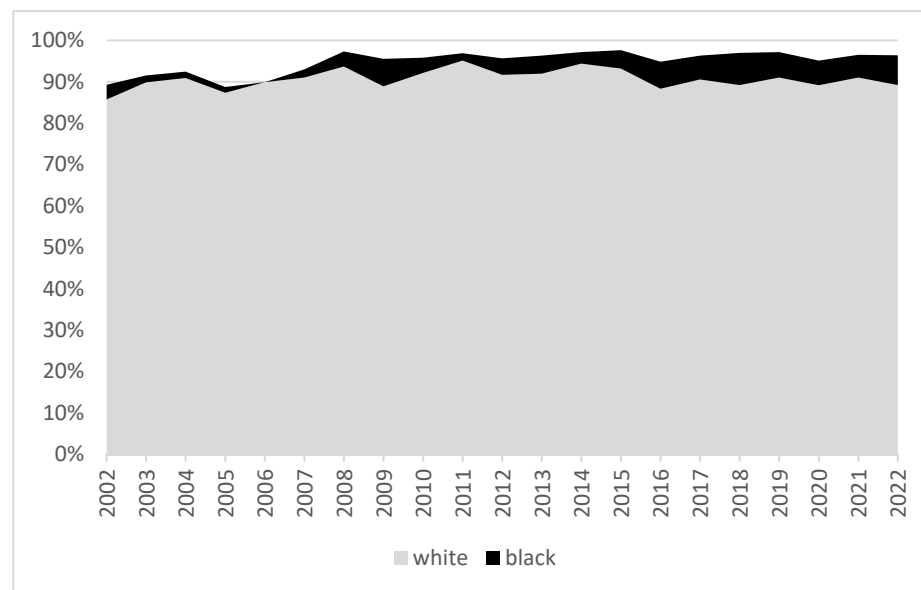
Despite the fact that, as stated above, colors besides black and white are used much less, there might still be some trends to identify regarding the increase or decrease of their

use during these past decades. For this reason, their values were normalized between the values 0% and 100% where 100% was the value of the time when the color was most popular. These normalized values are presented in Figure 17.



**Figure 17.** Basic RYB color usage by year (normalized) for (a) black and white, (b) warm colors, and (c) cool colors.

In order to explore the usage of empty space in the collected website instances over the years, the most prevalent color detected in every instance was considered the empty space color. Figure 18 presents an area chart that depicts how the use of white and black as empty space colors evolved in the past decades. It is noticeable that while the use of white remains mostly steady, the use of black displays a small but significant positive trend.



**Figure 18.** Usage of the colors white and black as empty space colors.

On some occasions, other basic RYB colors were detected as the empty space colors, though that percentage for each individual color was relatively small. Figure 19 presents

an area chart that depicts the use of other basic RYB colors as empty space colors over the past decades. An overall negative trend in the use of other colors is apparent, with a major drop in their usage around 2007 and 2008.

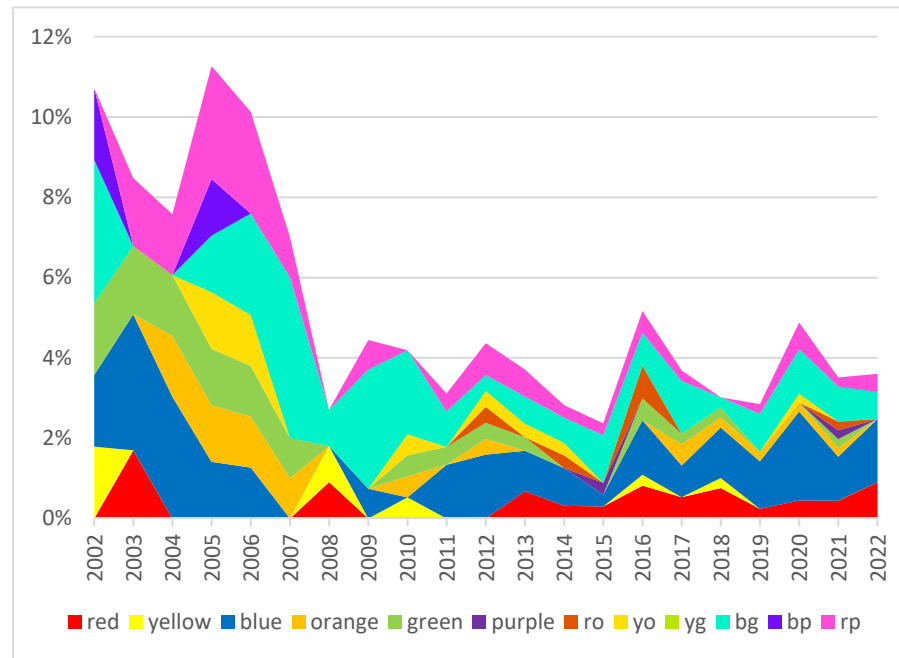


Figure 19. Usage of basic colors besides white and black as empty space colors.

Finally, as mentioned in the methodology section, the various color schemes that were identified in the collected website instances are presented in Figure 20. The monochromatic scheme, which uses one more basic color besides black and white, is prevalent throughout the studied period. At the same time there seem to be small positive trends in the usage of two complementary or two analogous colors. The use of three triadic colors is occasionally detectable but appears to be very limited throughout the past decades.

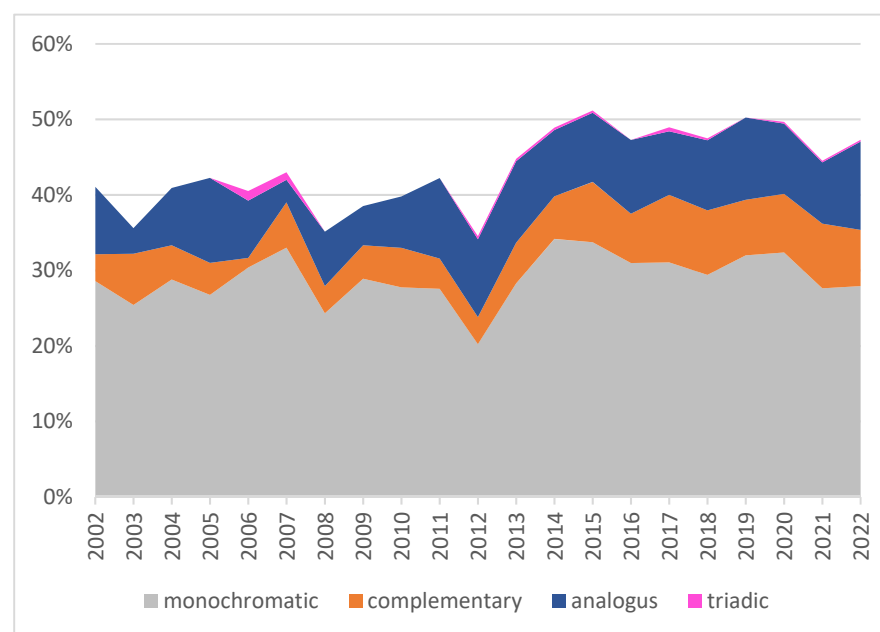


Figure 20. Usage of color schemes by year.

## 4. Discussion

### 4.1. RQ1: How Has the Integration of Semantic Web Technologies Progressed in the Last Decades? When and to What Extent Were Various Technologies Implemented?

As seen in Figure 11, the first signs of support for machine-readable content in the media outlets studied appeared in 2005. As Powers notes [25], the RSS2.0 specification was released in late 2002. From our findings, it appears that the practice of providing RSS2.0 feeds for public use started getting popular soon after. By 2010, almost half of the investigated website instances supported RSS.

Soon afterward, the adoption of Semantic HTML elements and Open Graph RDFa data begins. Back in the early 2010s, Fulanovic et al. [21] pointed out the importance of using the semantic elements instead of classes and ids to provide contextual information on websites and from our findings, it appears that this importance was acknowledged in the field of media outlets. In the same time period, the downward trajectory in traditional news media, alongside the rise of social media, as noted by Bajkiewicz et al. [33], dictated a shift from traditional media relations to a hybrid model, making most out of the Social Media environment. These facts support the extremely steep adoption curves that both HTML and OpenGraph displayed in Figure 11.

Within a few years, Semantic HTML elements gain presence in over 70% of website instances and by today almost 90% of website instances make use of at least some of them. In a similar style, more than 80% of modern websites are using Open Graph. Twitter Cards, which is also very closely connected to Social Media, follows a similar trend with Open Graph, just a couple of years later.

Schema.org data structures start getting identified as early as 2012 but did not start their steep climb in popularity until 2015. As noted by Muesel et al. [34], with schema.org's backing from major search engines, its adoption has been widespread. In a more recent study by Giannakoulouopoulos et al., the usage of schema.org is found to be a competent predictor for Web traffic based on popularity in art and culture-related media outlets [35]. Based on all of the above it is apparent that SWT adaptation in media outlets is high and it is safe to assume that higher visibility and content diffusion are the main motivations behind this.

### 4.2. RQ2: What Are the Trends in Website Aesthetics That Can Be Identified Concerning the Complexity of Web pages, the Usage of Graphics, and the Usage of Fluid or Responsive Designs?

As clearly displayed in Figure 12, complexity, both as measured by `<div>` elements and as measured by hyperlinks are linearly increasing over the passage of time. The rate of increase is higher in `<div>` tags than it is in hyperlinks, but they are both overall pretty similar curves. As mentioned above, according to Harper et al. [27] HTML structural complexity is related to how visual complexity is perceived by the website visitors. According to King et al. [8], high levels of visual design complexity will result in both more favorable user first impressions and increase the users' perception of both visual informativeness and cues for engagement. This indicates a strong motivation for media outlet websites to present the user with such complexity. It should be noted that other studies such as Chassy et al. [28] and Harper et al. [27] had contradictory findings, with visual complexity appearing to negatively impact aesthetic pleasure. The difference may lie with the focus, which in one case was on informativeness and engagement and in the other cases on aesthetic pleasure. Users may judge a visually complex site as informative while a visually simpler site as beautiful. In the case of online media outlets, the first case is more in line with the website's intended purpose. In a similar manner as displayed in Figure 13, most graphical elements present positive trends throughout the years, which in turn lead to higher design complexity, which according to King et al. [8] can lead to much coveted favorable first impressions concerning informativeness. An exception is the image map (`<map>`) graphical element. The difference between this element and the rest is that it does not adjust to fluid or responsive layout design since its dimensions are fixed. With the rise of mobile Internet, it is expected that its usage has plummeted. The video element

also diverts from the norm, because there appears to be a fall in its usage after 2017 which has only recently been reversed. More detailed investigation through further research might shed some light on this matter, but it is noteworthy that in April 2018 was when Chrome and other Chromium-based browsers changed their autoplay policy to not allow video autoplay, in order to minimize the incentive for ad blockers and reduce mobile data consumption [36].

A major drawback of increased complexity, both in terms of structure and in terms of graphical elements, is that a large website file size negatively affects the website's loading times which can have an adverse effect on SEO [37]. However as time goes by, such technical limitations are overcome through the development of faster networks and devices with higher processing power which ensure fast loading times in increasingly large file sizes.

In terms of fluid and responsive design, Figure 14 paints a clear picture, with table elements diminishing while mobile support is increasing alongside grid-based responsive frameworks. The simple approach of table elements with fluid widths was a good first step into multiple screen resolution support, but with the mobile Internet becoming more prevalent, more than that was required. It is symbolically significant that in Figure 14, the table element curve crosses the mobile scale curve sometime in 2014, which was the year that mobile Internet usage exceeded that of the PC for the first time [38]. The *mobile\_scale* variable reaches up to 90% in 2022, further reinforcing its significance. When it comes to specific frameworks, Bootstrap hovers above 30%, while Foundation is much lower. It is safe to assume that there are other ways to achieve repressiveness that our algorithm did not detect since there can be differences in framework keywords even from version to version of the same framework. Nevertheless, the rise in popularity of responsive web design tools is apparent from 2015 and onwards.

#### 4.3. RQ3: What Basic Colors and Coloring Schemes Are Prevalent in Website Homepages? Did They Change over the Years and Are There Consistent Trends That Can Be Inferred by Such Changes?

As depicted in Figure 15, the number of RYB basic colors besides black and white used in website instances over the past decades is slowly but steadily increasing. Despite that, website instances only using black, white, and shades of gray were still the relative majority in 2022. Other than that, using one or two additional colors were also popular choices, and these three categories together constituted over 80% of our sample throughout the recent decades.

From a usage perspective, as seen in Figure 16, white and black are, of course, the most used colors. Usage of the other basic RYB colors is much less prevalent since, as we established in Figure 15, very few are usually used in each website instance. The normalized color usage graphs in Figure 17 can be used to identify trends in color usage. White seems to be displaying significant stability with a very limited decline noticed in the latest years. On the other hand, usage of black is increasing. The graphs for other colors do not display a clear pattern and can be quite erratic on a case-by-case basis. From the warm colors, red and red-orange seem to display a positive trend, while from the cool colors purple displays a similar pattern. Previous research by Alberts and Van der Geest has tried to link specific color usage on the Web with trustworthiness, finding blue, green, and red to be most positively linked to user perception of trust [10]. Bonnardel et al. investigated the various colors that appeal to users and Web designers and concluded that blue and orange were considered the most appealing [11]. Both these studies took place in 2011 after which both blue and orange seem to present a positive trend in our findings too. On the other hand, the use of green, which was considered second most related to trustworthiness by Alberts [10], has been steadily declining in our findings and is, in terms of usage, the third least used color overall. The blue-green tertiary color though has been found to be one of the most popular colors just behind white, black, and blue. Overall, the trends regarding specific color usage presented in Figure 17 can be used to draw rather limited conclusions. As Swasty et al. [39] note, user responses to color vary based on different



demographic factors such as age and gender. Additionally, what is important is not the specific color used but successfully utilizing that color to build brand identity. Despite that, Talaei describes that emotional response to specific colors is part of human nature [12] and our findings confirm that there are indications that the use of specific colors is a conscious design choice, aiming to create appeal. Nevertheless, further research work is required to draw safer conclusions.

On the matter of empty space color, white is the most popular choice, with black being a very distant second choice as seen in Figure 18. As Eisfeld and Kristallovich [40] present in a recent study, the light-on-dark color scheme, which is based on black as an empty space color, has been increasingly popular and has ushered the coming of “Dark Mode” in applications and websites. The intent of such an approach is reduced eye strain, as overall screen time for individuals increases [40]. On the other hand, as seen in Figure 19, using colors besides black and white as empty space colors has declined. A major factor in this development might be the fact that modern human–computer interaction design principles request a standard minimum contrast ratio which should be extended as discussed by Ahamed et al. [41] to improve both luminance and clarity. When using a color with above 16% brightness or saturation (which were the limits of black and white in our study) this contrast ratio is rather harder to achieve. Hence the media outlets’ motivation to increase accessibility and usability might lead to the abandonment of using other basic colors as empty space colors.

Finally, Figure 20 presents the usage of the various color schemes. The general effort to produce visual complexity which can lead to improved first impressions from users [8], while at the same time maintaining color harmony, leads to the increase of complementary and analogous color schemes in the last decade. White [13], especially mentions that the use of complementary color schemes that evoke pleasure, can invoke positive attitudes towards advertisements and drive purchases. On the other hand, the triadic scheme still amounts to a very small portion of the website instances that were investigated in this research.

## 5. Conclusions

In this research, an innovative method was used to collect information from the HTML source code and homepage screenshots of a large number of websites, over a period of two decades, using data extraction techniques on archival data. The websites investigated were the top 1000 online media outlets based on Web traffic in Greece and included websites of both international media outlets and Greek national and local media outlets. The main goal of the study was to observe the course of these websites throughout the past decades, in regards to the adaptation of popular Semantic Web technologies and the aesthetic evolution of their interfaces, which included aspects concerning DOM structure and visual complexity, fluid, and responsive layout design techniques, and color usage and schemes.

The introduction of SWT in the websites was fast and extensive, with the main motivation behind it being the greater diffusion of media content. Structural and visual complexity displayed a steady but significant positive trend, aiming to achieve better first impressions while still maintaining performance across a plethora of devices. The rise of the mobile Internet guided the investigated websites to the adoption of responsive web design principles. An increase of visual complexity was also noted in the usage of colors, accompanied not only by an effort to better abide by the principles of accessibility, as established by the use of black as an empty space color but also by an effort to more closely adhere to color harmony through the use of color combinations.

The study’s sample is large but does present limitations, in the sense that the criteria for selection were popularity on the Greek Web. Focusing on websites popular in a different country might have presented different results due to cultural or other factors. That being said, many of the studied websites were international media outlets, which would be popular in most of the world. An additional limitation of the research can be found in its focus on websites with high traffic, which might be inclined to adopt current technologies and trends more rapidly. Finding a more varied sample of media outlets that would

include low traffic or niche outlets could provide an interesting contrast. In the future, this research can be expanded to different fields of online activity, beyond news and media, and attempt to find comparable results. Additionally, focusing on regions with a large cultural distance to Greece could lead to conclusions regarding the connection between cultural identity and aesthetic trends. Moving forward, we will focus our future work on collecting information regarding a vast array of websites from different fields, beyond news outlets, while simultaneously adapting our metrics to better identify regional aesthetic trends, in order to contrast their development to global trends.

The World Wide Web is a constantly evolving entity that is influenced both by the rise and fall of technologies and by the continuous evolution of human nature through cultural trends, global events, and globalization in general. Studies of the Web's past and its course through time can provide valuable knowledge, pertaining not only to the present but hopefully preparing us for the future. The advancements of the Semantic Web and the aesthetic evolution of user interfaces can be useful tools at the disposal of every online media outlet, both established and new, and can lead to the overall betterment of the undeniable services they provide.

**Author Contributions:** Conceptualization, A.L. and A.G.; data curation, A.L. and M.P. (Minas Pergantis); formal analysis, A.L., M.P. (Minas Pergantis), and M.P. (Michail Panagopoulos); investigation, A.L. and M.P. (Minas Pergantis); methodology, A.L., M.P. (Minas Pergantis), M.P. (Michail Panagopoulos), and A.G.; project administration, A.G.; resources, M.P. (Minas Pergantis) and A.G.; software, A.L. and M.P. (Minas Pergantis); supervision, M.P. (Michail Panagopoulos) and A.G.; validation, M.P. (Michail Panagopoulos) and A.G.; visualization, A.L.; writing—original draft, A.L. and M.P. (Minas Pergantis); writing—review and editing, A.L., M.P. (Minas Pergantis), and M.P. (Michail Panagopoulos). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available in Zenodo at (<https://doi.org/10.5281/zenodo.6624915>, accessed on 7 June 2022), reference number (10.5281/zenodo.6624915).

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

## References

1. Díaz Noci, J. A history of journalism on the Internet: A state of the art and some methodological trends. *RIHC Rev. Int. Hist. Comun.* **2013**, *1*, 253–272. [[CrossRef](#)]
2. Karlsson, M.; Holt, K. Journalism on the Web. In *Oxford Research Encyclopedia of Communication*; Oxford University Press: Oxford, UK, 2016.
3. Deuze, M. Journalism and the Web: An analysis of skills and standards in an online environment. *Gazette* **1999**, *61*, 373–390. [[CrossRef](#)]
4. Berners-Lee, T.; Hendler, J.; Lassila, O. The semantic web. *Sci. Am.* **2001**, *284*, 34–43. [[CrossRef](#)]
5. Fernandez, N.; Blazquez, J.M.; Fisteus, J.A.; Sanchez, L.; Sintek, M.; Bernardi, A.; Fuentes, M.; Marrara, A.; Ben-Asher, Z. News: Bringing semantic web technologies into news agencies. In Proceedings of the International Semantic Web Conference, Athens, GA, USA, 5–9 September 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 778–791.
6. Heravi, B.R.; McGinnis, J. Introducing social semantic journalism. *J. Media Innov.* **2015**, *2*, 131–140. [[CrossRef](#)]
7. Wu, O.; Han, M. Screenshot-based color compatibility assessment and transfer for Web pages. *Multimed. Tools Appl.* **2018**, *77*, 6671–6698. [[CrossRef](#)]
8. King, A.J.; Lazard, A.J.; White, S.R. The influence of visual complexity on initial user impressions: Testing the persuasive model of web design. *Behav. Inf. Technol.* **2020**, *39*, 497–510. [[CrossRef](#)]
9. Stoeva, M. Evolution of Website Layout. In Proceedings of the Techniques Anniversary International Scientific Conference “Computer Technologies and Applications”, Pamporovo, Bulgaria, 15–17 September 2021.
10. Alberts, W.; Van der Geest, T. Color Matters: Color as Trustworthiness Cue in Web Sites. *Tech. Commun.* **2011**, *58*, 149–160.
11. Bonnardel, N.; Piolat, A.; Le Bigot, L. The impact of colour on Website appeal and users' cognitive processes. *Displays* **2011**, *32*, 69–80. [[CrossRef](#)]
12. Talaei, M. Study of human reactions than color and its effects on advertising. *Int. J. Account. Res.* **2013**, *42*, 1–9. [[CrossRef](#)]

13. White, A.E.R. Complementary Colors and Consumer Behavior: Emotional Affect, Attitude, and Purchase Intention in the Context of Web Banner Advertisements. Ph.D. Thesis, Universidade Nova de Lisboa, Caparica, Portugal, 2018. Available online: <http://hdl.handle.net/10362/52273> (accessed on 7 June 2022).
14. Brügger, N. The archived website and website philology. *Nord. Rev.* **2008**, *29*, 155–175. [CrossRef]
15. Gomes, D.; Miranda, J.; Costa, M. A survey on web archiving initiatives. In Proceedings of the International Conference on Theory and Practice of Digital Libraries, Berlin, Germany, 25–29 September 2011; Springer: Berlin/Heidelberg, Germany.
16. Internet Archive. About the Internet Archive. Available online: <https://archive.org/about/> (accessed on 1 June 2022).
17. SimilarWeb. We Are the Official Measure of the Digital World. Available online: <https://www.similarweb.com/corp/about/> (accessed on 1 June 2022).
18. Koehler, W. Web page change and persistence—A four-year longitudinal study. *J. Am. Soc. Inf. Sci. Technol.* **2002**, *53*, 162–171. [CrossRef]
19. Lamprogeorgos, A.; Pergantis, M.; Giannakouloupoulos, A. A methodological guide to gathering archival data related to website aesthetics. In Proceedings of the 4th International Conference Digital Culture & AudioVisual Challenges, Corfu, Greek, 13–14 May 2022. *Under publication*.
20. HTML Semantic Elements. Available online: [https://www.w3schools.com/html/html5\\_semantic\\_elements.asp](https://www.w3schools.com/html/html5_semantic_elements.asp) (accessed on 1 June 2022).
21. Fulanovic, B.; Kucak, D.; Djambic, G. Structuring documents with new HTML5 semantic elements. In Proceedings of the 23rd DAAAM International Symposium on Intelligent Manufacturing and Automation, Zadar, Croatia, 24–27 October 2012; Volume 2, pp. 723–726.
22. The Open Graph Protocol. Available online: <https://ogp.me/> (accessed on 1 June 2022).
23. About Twitter Cards. Available online: <https://developer.twitter.com/en/docs/twitter-for-websites/cards/overview/abouts-cards> (accessed on 1 June 2022).
24. Infante-Moro, A.; Zavate, A.; Infante-Moro, J.C. The influence/impact of Semantic Web technologies on Social Media. *Int. J. Inf. Syst. Softw. Eng. Big Co.* **2015**, *2*, 18–30.
25. Powers, S. *Practical RDF*; O'Reilly Media, Inc.: Cambridge, MA, USA, 2003; pp. 10, 254.
26. Mika, P. On schema.org and why it matters for the web. *IEEE Internet Comput.* **2015**, *19*, 52–55. [CrossRef]
27. Harper, S.; Jay, C.; Michailidou, E.; Quan, H. Analysing the visual complexity of web pages using document structure. *Behav. Inf. Technol.* **2013**, *32*, 491–502. [CrossRef]
28. Chassy, P.; Fitzpatrick, J.V.; Jones, A.J.; Pennington, G. Complexity and aesthetic pleasure in websites: An eye tracking study. *J. Interact. Sci.* **2017**, *5*, 3. [CrossRef]
29. Bootstrap Team. Bootstrap—The Most Popular HTML, CSS and JS Library in the World. Available online: <https://getbootstrap.com/> (accessed on 1 June 2022).
30. ZURB. Foundation—The Most Advanced Responsive Front-End Framework in the World. Available online: <https://get.foundation/> (accessed on 1 June 2022).
31. Art of the Web. PHP: Extracting Colours from an Image. Available online: <https://www.the-art-of-web.com/php/extract-image-color/> (accessed on 1 June 2022).
32. Gage, J. *Color and Meaning: Art, Science, and Symbolism*; University of California Press: Oakland, CA, USA, 1999.
33. Bajkiewicz, T.E.; Kraus, J.J.; Hong, S.Y. The impact of newsroom changes and the rise of social media on the practice of media relations. *Public Relat. Rev.* **2011**, *37*, 329–331. [CrossRef]
34. Meusel, R.; Bizer, C.; Paulheim, H. A web-scale study of the adoption and evolution of the schema.org vocabulary over time. In Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics, New York, NY, USA, 13–15 July 2015; pp. 1–11.
35. Giannakouloupoulos, A.; Pergantis, M.; Konstantinou, N.; Kouretsis, A.; Lamprogeorgos, A.; Varlamis, I. Estimation on the Importance of Semantic Web Integration for Art and Culture Related Online Media Outlets. *Future Internet* **2022**, *14*, 36. [CrossRef]
36. Beaufort, F. Autoplay Policy in Chrome. 2017. Available online: <https://developer.chrome.com/blog/autoplay/> (accessed on 1 June 2022).
37. Chotikitpat, K.; Nilsook, P.; Sodsee, S. Techniques for improving website rankings with search engine optimization (SEO). *Adv. Sci. Lett.* **2015**, *21*, 3219–3224. [CrossRef]
38. Murtagh, R. Mobile Now Exceeds PC: The Biggest Shift Since the Internet Began. *Search Engine Watch* **2014**. Available online: <https://www.searchenginewatch.com/2014/07/08/mobile-now-exceeds-pc-the-biggest-shift-since-the-Internet-began/> (accessed on 1 June 2022).
39. Swasty, W.; Adriyanto, A.R. Does color matter on web user interface design. *CommIT (Commun. Inf. Technol.) J.* **2017**, *11*, 17–24. [CrossRef]
40. Eisfeld, H.; Kristallovich, F. The rise of dark mode: A qualitative study of an emerging user interface design trend. *Jönköping* **2020**. Available online: <http://hj.diva-portal.org/smash/get/diva2:1464394/FULLTEXT01.pdf> (accessed on 1 June 2022).
41. Ahamed, M.; Bakar, Z.; Yafooz, W. The Impact of Web Contents Color Contrast on Human Psychology in the Lens of HCI. *Int. J. Inf. Technol. Comput. Sci.* **2019**, *11*, 27–33. [CrossRef]