

Audiovisual Art Event Classification and Outreach Based on Web Extracted Data

Andreas Giannakouloupoulos^{}, Minas Pergantis^{}, Aristeidis Lamprogeorgos^{},
Stella Lampoura^{}

Department of Audio and Visual Arts, Ionian University, Corfu, Greece

Email: agiannak@ionio.gr, mperg@ionio.gr

How to cite this paper: Giannakouloupoulos, A., Pergantis, M., Lamprogeorgos, A. and Lampoura, S. (2025) Audiovisual Art Event Classification and Outreach Based on Web Extracted Data. *Journal of Software Engineering and Applications*, 18, 24-43.

<https://doi.org/10.4236/jsea.2025.181002>

Received: December 20, 2024

Accepted: January 14, 2025

Published: January 17, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0).

<http://creativecommons.org/licenses/by-nc/4.0/>



Open Access

Abstract

The World Wide Web provides a wealth of information about everything, including contemporary audio and visual art events, which are discussed on media outlets, blogs, and specialized websites alike. This information may become a robust source of real-world data, which may form the basis of an objective data-driven analysis. In this study, a methodology for collecting information about audio and visual art events in an automated manner from a large array of websites is presented in detail. This process uses cutting edge Semantic Web, Web Search and Generative AI technologies to convert website documents into a collection of structured data. The value of the methodology is demonstrated by creating a large dataset concerning audiovisual events in Greece. The collected information includes event characteristics, estimated metrics based on their text descriptions, outreach metrics based on the media that reported them, and a multi-layered classification of these events based on their type, subjects and methods used. This dataset is openly provided to the general and academic public through a Web application. Moreover, each event's outreach is evaluated using these quantitative metrics, the results are analyzed with an emphasis on classification popularity and useful conclusions are drawn concerning the importance of artistic subjects, methods, and media.

Keywords

Web Data Extraction, Art Events, Classification, Artistic Outreach, Online Media

1. Introduction and State-of-the-Art

Over the years, the World Wide Web has emerged as the primary means of communication and information sharing, resulting in significant shifts in how people

understand the process of consuming information [1]. In this contemporary landscape, information about artistic works and cultural heritage artifacts has also migrated into the digital realm through digitization [2]. To provide users with an engaging experience when interacting with content related to art and culture, it is crucial to comprehend their behavioral patterns [3] and to utilize that insight in crafting a user experience (UX) that caters to the distinct needs of both the users [4] and the content itself. This necessity is particularly heightened in outlets presenting vast amounts of information, as the user interface (UI) is an essential instrument to mitigate user confusion [5] and enhance engagement by prioritizing simplicity and familiarity.

It is safe to say that a key aspect of sharing cultural events, including various audiovisual art projects, is their visibility on the Internet [6]. Online media platforms, cultural websites, blogs, social media, and other digital repositories offer abundant information about art events. However, news coverage available on the web is frequently both scattered and short-lived, influenced by the differing interests of traditional media and user-generated content [7] [8], as well as the popularity of algorithm-driven news distribution systems [9]. To gather more meaningful insights into the traits of modern audiovisual art projects, it is essential to extract information from a variety of sources. Additionally, the process of pinpointing and gathering details regarding a specific event from different websites can result in a well-informed assessment of its potential reach, which may subsequently affect its cultural significance [10].

In order to collect large-scale information from the Web in an automated manner, it is essential to programmatically analyze the structure of Web pages and their metadata, as they appear in social media data-graphs or Semantic Web integrations [11]. Additionally, this data must be recorded in a manner that takes advantage of the benefits of Semantic Web ontologies and data models. The Semantic Web arose as an effort to enhance how information is structured online to improve its ability to convey meanings in a manner that digital systems can interpret. This process includes the systematic arrangement of data into standardized, subject-oriented hierarchies known as ontologies [12]. Ontologies serve as a means of comprehending and organizing web information and lie at the heart of the Semantic Web, facilitating the development of a versatile digital space that acts not just as a collection of relevant data but as a dynamic, intelligent, and efficient environment with new opportunities for advanced applications and services [13]. Although computers do not “understand” information like humans do, they can efficiently relate and manage it so that the outcome becomes meaningful for human users [14]. Scholars in Web science are continuously working on research and technological advancements to actualize the Semantic Web concept [15]. For instance, the RDF/XML standard has shown to be challenging to apply in scenarios with high complexity [16], whereas newer technologies such as N-Triplets [17], Turtle [18], and JSON-LD [19] offer more readable structures, enabling better grammar and smoother mapping from XML syntax to RDF. As a result, the

technologies capable of realizing the essential Semantic Web are evolving and will continue to develop, moving closer to achieving semantic understanding of data by machines.

Going one step further, Generative Artificial Intelligence (GenAI) can be used to both analyze html documents and derive structured information from their texts. AI technologies have a long history of successful application on text structuring and summarization [20] which has become even more powerful with the advances in AI and the introduction of the transformer architecture, which gave birth to GenAI as we know it [21].

Mining information from the Web may lead to large objective datasets which are quintessential in statistical analysis. In recent years, there has been a growing interest among researchers around big data and computational engineering. Big data is defined as an extensive collection of datasets that conventional data processing systems cannot analyze; instead, they demand innovative automated technologies for tasks like capturing, storing, distributing, managing, and analyzing [22]. Online news items are an example of such datasets. This vast and varied information is unmanageable through manual processing, which involves breaking it down into sentences or words to analyze and extract valuable insights [23]. Additionally, significant attention from researchers has been directed toward the application of big data, highlighting its role as a crucial indicator of a company's efficiency and productivity [24]. This use of big data has also found its way into social sciences and culture in the form of cultural analytics which may be defined as "the analysis of massive cultural data sets and flows using computational and visualization techniques" [25].

In this study, modern innovative technologies are utilized to create a large objective dataset from the Web concerning audiovisual art events as they are presented by online media outlets and other related websites. Via the use of programmatic Web Search, Semantic Web analysis and the power of GenAI structured information is collected about the various aspects of an event's online presence. This information is quantified and processed to derive objective measurements, which are used for statistical analysis. This article presents the detailed methodology of both data collection and an analysis of real-world data regarding contemporary works of audio and visual art in Greece. The collected data are openly provided to the general public through an innovative Web application which focuses on user experience and functionality [26] available at <https://repo.artdata.gr/> and through an extended application programming interface. The analysis focuses on the classification of the collected events based on the event type, the methods and techniques used in the event as well as the events' thematic elements. By investigating both the popularity of specific classification tags and the popularity and perceived cultural outreach of individual events described by these tags, the study demonstrates how objective Web data may be collected, analyzed and used to reach significant conclusions about the importance of various artistic subjects, methods and media in contemporary culture. The following research questions

concerning the Greek audiovisual event landscape were established as a means to guide the research direction:

RQ1. What are the most popular classifications of audiovisual art events regarding event type, methods, and subjects?

RQ2. What are the classifications of audiovisual art events regarding event type, methods and subjects that are most influential based on event outreach index?

2. Methodology

The algorithm for data extraction is made up of three primary components. The initial phase involves searching through media websites, which, after being processed, results in a collection of audiovisual events. The second and most important component includes an iterative process that seeks additional publications related to each of these events on the internet, transforming the content of these publications from code and free text found on any webpage into a graph of structured data that is integrated into the recorded information about the event in the system's database. Additionally, during this phase, publications that are extracted but do not relate to an event contribute to the formation of new events for further investigation. Lastly, the third component investigates the websites where the publications retrieved during the event inquiry originated and gathers information from them.

2.1. Event Data Web Extraction Algorithm

Before classifying and analyzing the outreach of an audiovisual event, it is important to generate a record highlighting its characteristics. The data extraction algorithm begins by collecting initial events, discovering them from means that distribute these events online. It starts from specific web pages that publish art and culture events, programmatically extracting URLs of relevant publications through web mining. Once a list of publication URLs is generated, the algorithm retrieves the content of each individual web page. The information is then integrated into a data graph that fits the system's data model, utilizing semantic web technologies to identify web page structure. GenAI methods are employed to analyze the texts and to extract structured data and summaries, which are stored in the system's relational database. These structured data include classification tags based on event type, themes, and media. The next step involves inferring events from these publications. The algorithm compares the data with existing event records to determine if a publication is related to an already documented event. If a match is found, the publication information is integrated into the event details. If not, a new event record is created, increasing the total number of recorded events, and preparing the analysis of the outreach of each event.

The discovery of the full range of a specific event's classification tags and the analysis of its Web outreach starts by utilizing a search engine API to locate more publications related to the event's title and location. The API provides distinct URLs that contain pertinent content, which will be the basis for further examination.

Following this, the algorithm gathers additional information from these web pages. By employing GenAI, it determines if the content pertains to an audiovisual event. If it confirms this, the text is converted into a structured data graph which includes classification tags and recorded in the database as a new publication entry. The newly added publication is subsequently assessed for its relevance to the original event using spatial and temporal derived from the event's title and description. If it is deemed relevant, the information and classification tags are integrated into the event; if not, they are added to the database to be investigated later. This procedure is repeated for every event in the database, and upon completion, a final step of publication comparison and media exploration is carried out.

Every publication that results from the search for related materials using the search engine API, as outlined above, comes from a distinct website that might not be part of the original list of websites from which the initial publications were obtained. In this scenario, the algorithm retrieves the main web page of that website and looks for information regarding the media outlet, using the semantic data supplied by the web page. The media exploration process establishes a fundamental record of the content provider in the system's database, which is then further enhanced with details about the outlet by human editors. This additional data includes information on its popularity and social media presence, which is valuable for exploring the extent of event dissemination.

2.2. Outreach Analysis Algorithm

To analyze the extensive data gathered and stored in the system's database, it was essential to quantify the distinct characteristics associated with the impact of each event. The intrinsic characteristics of the events, which can be measured directly, played a crucial role in this process, such as the total number of publications related to each event, or the count of contributors involved. Furthermore, by examining the textual content of these publications, quantitative assessments concerning the extent of each production and the target audience it addresses were derived with the assistance of GenAI. Concurrently, attention was also directed toward the attributes of the various media outlets that disseminate information about each event, particularly focusing on quantitative measures of their popularity, like website traffic statistics and social media presence size. Lastly, the popularity of the tags used to describe and categorize each event was assessed by considering the number of other collected events that share similar classification tags.

A key component of the algorithm designed to estimate the dissemination of different audiovisual art events was assessing the corresponding popularity of the online media outlets and websites that covered these events. This popularity was determined by the online visibility of the websites, which was characterized by their overall monthly traffic, backlinks, organic search rankings, ranking status in Greece, and their domain authority. Moreover, the presence of the media on various social media platforms and the size of their follower or subscriber counts on official accounts was also investigated. The following attributes were used in these

estimates, each with a different weight:

- Web Popularity
- Total Visits, weight 5
- Backlinks, weight 4
- Organic Results, weight 3.5
- Ranking in Greece, weight 2
- Domain Authority, weight 1.5
- Social Media Popularity
- YouTube, weight 5
- Facebook, weight 4.5
- X (Twitter), weight 3.5
- TikTok, weight 3
- Instagram, weight 1.5
- LinkedIn, weight 1

In terms of Web popularity, the total number of visitors to the website was considered the most crucial factor and therefore received the highest weight. Backlinks serve as a significant indicator of how integrated the website is within a broader informational network, and for this reason, it was also assigned a substantial weight. This was succeeded by the quantity of organic results, which pertain solely to visits originating from search engines. Lastly, both the Greek ranking and domain authority can reflect a domain's local or topic-specific significance, and for this reason, they were included as well, though with lesser weights. All the metrics mentioned were sourced through third-party services like Similar Web and Neil Patel.

Regarding social media popularity, the platforms chosen were Facebook, YouTube, Instagram, TikTok, X (Twitter), and LinkedIn. The first four are among the most visited according to SimilarWeb, while X (Twitter) was selected for its crucial role in spreading news and LinkedIn for its relevance in the business sector. Platforms that primarily focus on messaging, such as WhatsApp, WeChat, and Messenger, or those with limited presence in Greece, like eastern platforms (Douyin, Baidu, etc.), were excluded. YouTube received the highest rating due to its significance in delivering audiovisual content, followed closely by the well-established and widely used Facebook. X (Twitter) ranked next in importance because of its function as a news distributor. Although TikTok's audiovisual characteristics are significant, its narrower audience base resulted in a moderate score. Lastly, Instagram, with its restrictions on URL linking, and LinkedIn, which is focused on commerce, are also notable but less so than the others.

Moving from media outlet popularity to specific event outreach, a series of measurements were devised to quantitatively describe each individual audiovisual art event. These metrics included the weighted harmonic mean of the Web popularity and the Social Media popularity of each website that reported on the event, the total number of publications discovered during the web search process, the total number of contributors identified from the event publications' texts, the

intended audience size and the estimated production cost as inferred with the help of GenAI through the textual descriptions of the event on the various media outlets and finally the harmonically weighted sum of events that shared the same classification tags as the investigated event based on type, methods and subject.

The various metrics used for both the estimation of media outlet popularity and event outreach were normalized using logarithmic and z-score methods, depending on the number of outliers detected in each metric. The total popularity indexes were calculated using weighted means which gave increasing importance to the various metrics in accordance with the order they are presented in the list above. Finally, all the event related metrics were also consolidated into a singular event outreach metric using an equivalent weighted average methodology. The different quantitative metrics for each event and their weights are as follows:

- Web Popularity Harmonic Average, weight 5
- Social Popularity Harmonic Average, weight 4.5
- Publication Count, weight 3.5
- Estimated Audience Size, weight 2
- Estimated Production Cost, weight 2
- Contributor Count, weight 1
- Type Popularity Aggregate, weight 1
- Method Popularity Aggregate, weight 1
- Subject Popularity Aggregate, weight 1

Popularity through Web and social media, which involves various objective characteristics of the different media outlets, was identified as the most critical aspect, closely followed by the number of publications, as it serves as a valuable objective indicator of outreach. The frequency of each classification tag was assessed in a way that ensures that overall the various classifications' significance was considered moderately important. The values derived from publication texts were rated just slightly lower, as they provide additional insights, but may be subjective or inconsistent due to the nature of GenAI. Lastly, the metric with the least weight was the number of contributors, which offers some intrinsic value in reflecting the scale of an event, but is less reliable since media outlets often fail to report every contributor.

This singular index can be used as a strong indicator of individual event popularity to draw conclusions regarding the relationship between the various classification tags, their own popularity as well as the popularity of the events they describe.

Figure 1 presents a detailed flowchart of the outreach analysis algorithm.

3. Results

During a data collection period that took place in two separate phases in the Spring and Summer of 2024 over 16,500 different audiovisual art events in Greece were discovered. These events were reported in over 30 k unique webpages from over 2.4 k different media outlets and websites. The generation of classification

tags based on the texts of these discovered publications lead to over 1000 different tags describing the type of the events, over 1700 different tags describing the methods and techniques used in the events and over 4500 different tags describing the events' themes and subject matter. **Table 1** presents a summary of the collected information.

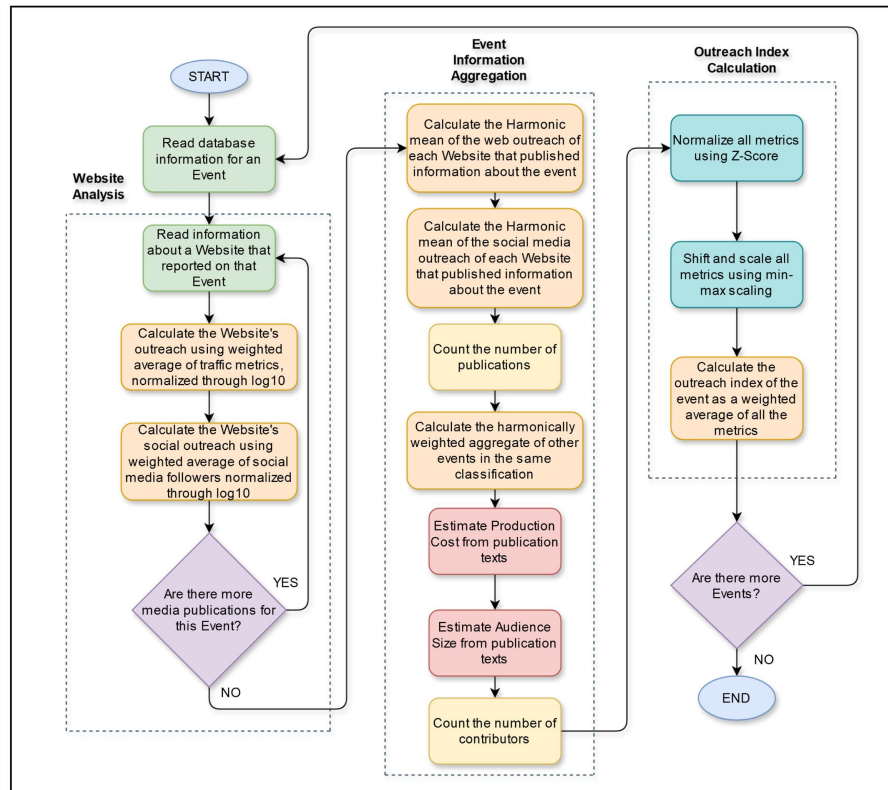


Figure 1. Outreach analysis algorithm flowchart.

Table 1. Summary of the web extraced information.

	# of occurrences
Events	16,501
Publications	30,803
Media	2456
Type tags	1149
Method tags	1775
Subject tags	4698

In order to gain insights into the popularity of the various classification tags the total sum of events which were described using one such tag was calculated. This led to a ranking based on how plentiful were the events which belonged to that specific tag. **Table 2** presents the top 20 most popular tags describing an event's

type. Since the data extraction process involved Greek events and publications mainly in the Greek language, the tag names have been translated into English for the reader's benefit.

Table 2. Top 20 most popular event type tags.

Ranking	Tag name	# of occurrences
1	Music	8247
2	Concert	4922
3	Culture	4397
4	Theater	4155
5	Cinema	1151
6	Festival	1002
7	Art	921
8	Comedy	767
9	Exhibition	680
10	Classical Music	663
11	Visuals	650
12	Dance	647
13	Documentary	502
14	Live Performance	494
15	Cultural Event	488
16	History	469
17	Performance	435
18	Feature	370
19	Visual Arts	326
20	Drama	321

The distribution of events over these tags is presented more clearly in **Figure 2** in the form of a pie chart. According to the chart music related events are dominant in the landscape of audiovisual art events, with the first most popular tags “music” and “concert” culminating in over 40% of the events. The performative arts follow with “theater” and “cinema” being in the top 5 most popular tags alongside the more generic tag of “culture”. The top 20 is populated with other general types of art events such as “exhibitions”, “visual art” and “dance”, but also includes more specific subcategories such as “comedy”, “classical music” and “drama”.

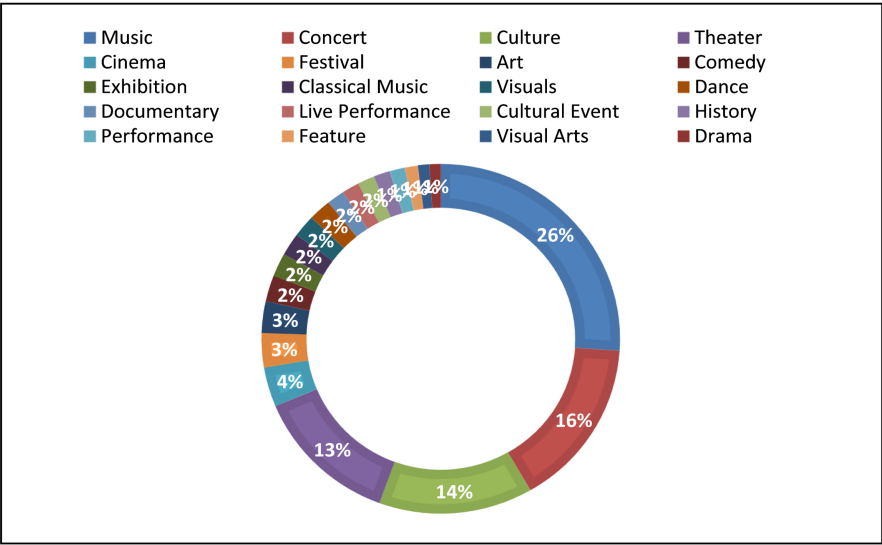


Figure 2. Pie chart of the distribution of type tags.

Going beyond simple event type categorization, the top 20 tags describing the means, methods and techniques used during the various audiovisual art events are presented in **Table 3**.

Table 3. Top 20 most popular event method tags.

Ranking	Tag name	# of occurrences
1	Live Music	4650
2	Theatrical Performance	2461
3	Music	2014
4	Direction	1668
5	Acting	1461
6	Dance	1433
7	Concert	1382
8	Vocals	697
9	Song Performance	673
10	Piano	646
11	Painting	646
12	Singing	558
13	Narration	555
14	Musical Performance	549
15	Guitar	523

Continued

16	Dramaturgy	480
17	Choir	478
18	Discussion	468
19	Drums	450
20	Speech	384

How the various tags relating to audiovisual art methods are distributed over the discovered events is presented more clearly in **Figure 3** in the form of a pie chart.

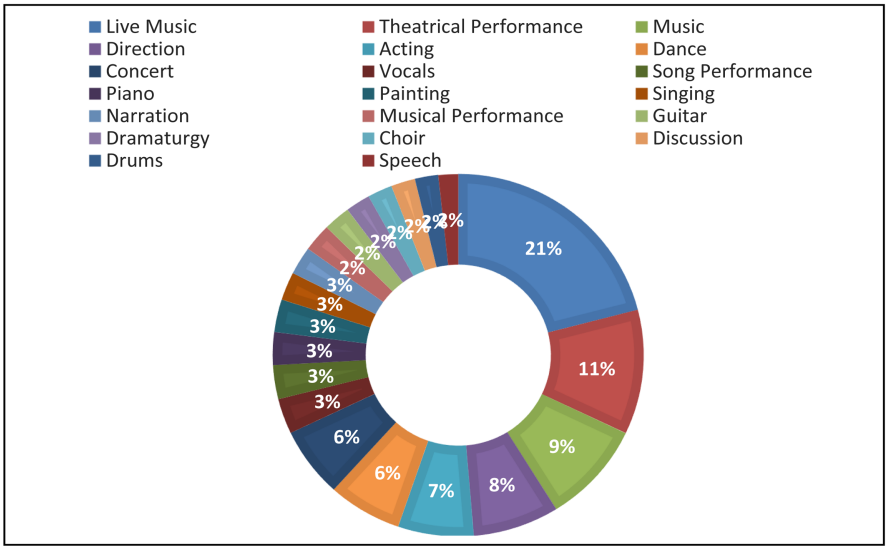


Figure 3. Pie chart of the distribution of method tags.

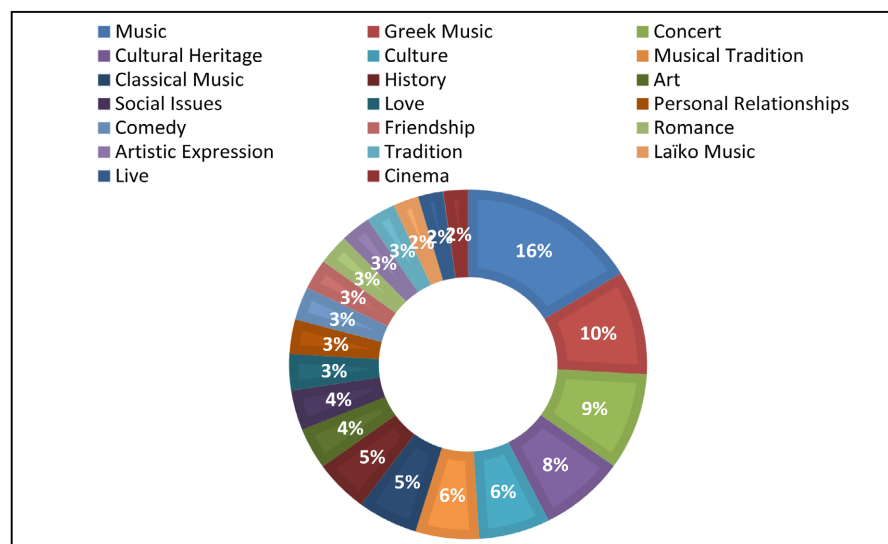
Continuing the trend observed in **Figure 2**, the more musical methods are dominant with “live music” getting first place with over 20% of the events and accompanied by other related tags such as “music”, “concert”, “vocals” and more. The performative art methods and means also play a dominant role with “theatrical performance”, “direction” and “acting” all appearing in the top 5. The chart also includes the performance of various musical instruments such as “piano”, “guitar” and “drums”, as well as methods of storytelling, including “dance”, “dramaturgy”, “narration” and “speech”. “Painting” is also present in the top 20 as a representative of the fine arts.

Finally, the third category of classification tags included the themes and subjects of the events. **Table 4** presents the top 20 most popular subject tags and the number of events described by each tag.

The distribution of the above tags can be presented in a more comprehensive manner through the pie chart available in **Figure 4**.

Table 4. Top 20 most popular event subject tags.

Ranking	Tag name	# of occurrences
1	Music	2559
2	Greek Music	1515
3	Concert	1406
4	Cultural Heritage	1211
5	Culture	1010
6	Musical Tradition	905
7	Classical Music	855
8	History	768
9	Art	600
10	Social Issues	582
11	Love	526
12	Personal Relationships	501
13	Comedy	482
14	Friendship	441
15	Romance	437
16	Artistic Expression	429
17	Tradition	416
18	Folk Music	364
19	Live	354
20	Cinema	354

**Figure 4.** Pie chart of the distribution of subject tags.

As presented in **Figure 4** the importance of musical events is conveyed by the multitude of musical styles that are dominant in the top 20 subject tags. Going beyond the very general “music” and “concert” we get “Greek music”, “musical tradition”, “classical music” and “Laïko music” (which is a very prevalent blend of folk and popular music in Greece) [27]. “Culture” and “cultural heritage” are also popular tags and alongside “history” and “tradition” indicate a tendency for many events to focus thematically on Greece’s rich cultural past. Themes and subjects popular in performative art are also included in the chart with “social issues”, “love”, “romance”, “friendship” and “personal relationships” all being prevalent in the top 20 most popular subject tags.

Tables 2-4 and **Figures 2-4** give an interesting glimpse on the landscape of audiovisual arts in contemporary Greece, based on the popularity of the various classification tags that were mined from texts reporting on audiovisual art events on the Web. These real-world data represent an objective reality as reported through the web. Music, performative art, and other forms of expression are all represented in various measurable degrees. But the number of events for each specific classification tag cannot be the only indicator. The difference in popularity, outreach and cultural impact between various events also plays a significant role.

Out of the total 16,501 events a total of 2685 were fully investigated in terms of outreach using the web search algorithm to collect exhaustive information about them as described in the methodology section and using the outreach analysis algorithm to calculate their various outreach metrics. The culmination of these outreach measurements in one singular index allowed us to gauge the popularity of each event. For each classification tag that appeared in at least ten different fully investigated events an average outreach index was calculated. The threshold of ten events was established in order to have a representative sample of events before calculating the average. **Figure 5** presents the top 20 more impactful event type as a bar chart of the average outreach index of each type.

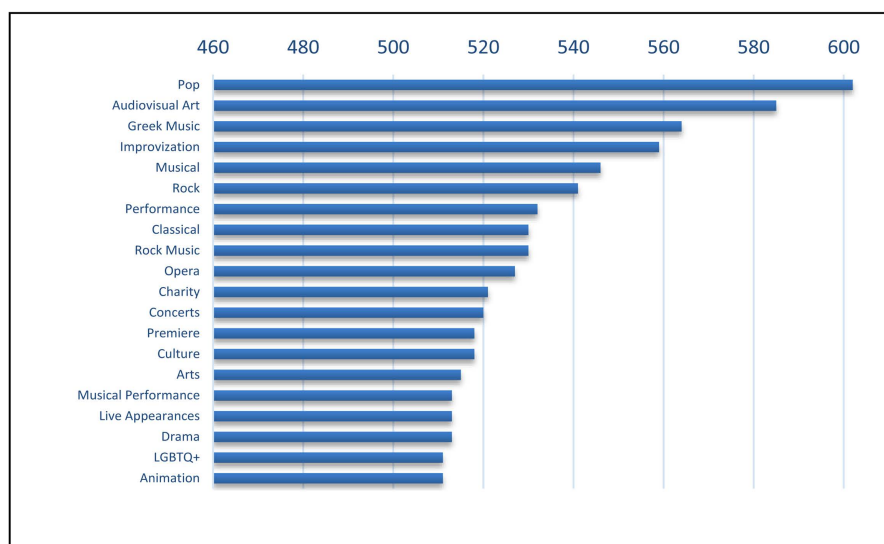


Figure 5. Bar chart of the type tags with the highest average outreach index.

Through the bar chart of the top 20 type tags with the highest average outreach index a series of highly impactful event type subcategories are established. Musical event genres such as “pop”, “rock”, “classical”, “opera” and “Greek music” all get higher places than the wider general tags like “music” got, indicating a higher outreach of these specific genres. Additionally, “audiovisual art” as a tag makes a strong appearance indicating a higher outreach for events with a clearer artistic identity. At the same time niche types like “improvisation” or “animation” make an appearance, and this is rounded up by some more general themes like “charity” or “LGBTQ+” issue related events.

Moving on to the audiovisual event classification tags for methods, techniques and media, **Figure 6** presents the top 20 tags with the highest average outreach index as a bar chart.

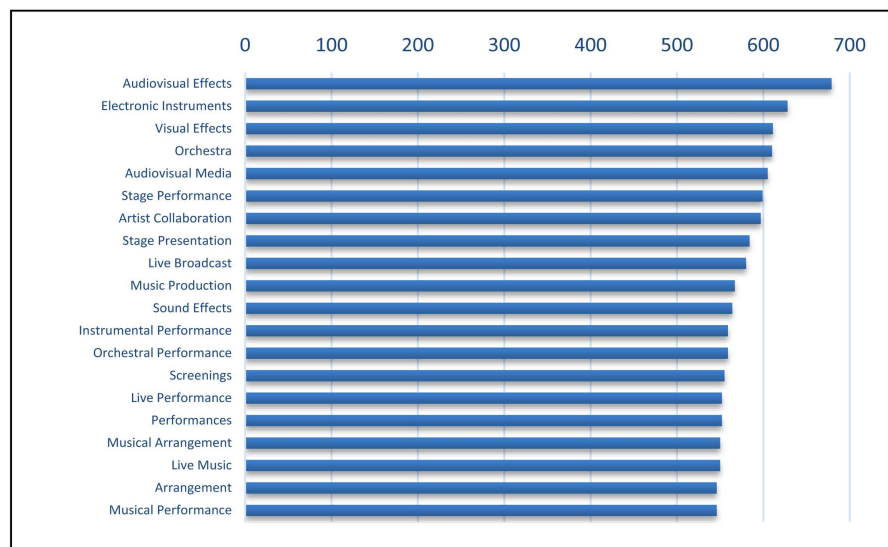


Figure 6. Bar chart of the method tags with the highest average outreach index.

As made clear by the bar chart, more technical presentation methods and techniques utilizing “audiovisual effects”, “visual effects”, “audiovisual media” and “electronic instruments” present a high average outreach index. Moreover, general technical aspects of an event are also highly important with “sound effects”, “stage presentation”, “musical arrangement” and “instrumental performance” making an appearance in the top spots. The high production values related to the use of modern technology and high technical implementation appear to have a big impact on the outreach of the events.

Finally, through **Figure 7**, we take a glimpse into the top 20 event classification tags for subjects and themes with the highest average outreach index in the form of a bar chart.

On the matter of subject classification tags, seasonal or annual themes have the highest average outreach index. The annual “Eurovision” song contest with its multitude of preliminary, semifinal, and final events appears in the top spot, followed by “New Year’s”, “celebration”, “anniversary” and “anniversary concert” in

the top 20, indicating the high outreach of seasonal events. The tag “Maria Callas” which also appears on the top 20 list is related to a multitude of events in tribute to 100 years from the artist’s birth, which also adds to the importance of anniversary related events. Music themes revolving around the “artist collaboration”, the “artistic collaboration” and “music career” point to a high outreach for events that have added significance due to scope or multitude of participants, while popular musical genres such as “pop”, “pop music” and “Greek songs” also allow for high outreach. Some tags related to “Greek culture”, such as “Greek tradition”, and “musical heritage” also make an appearance.

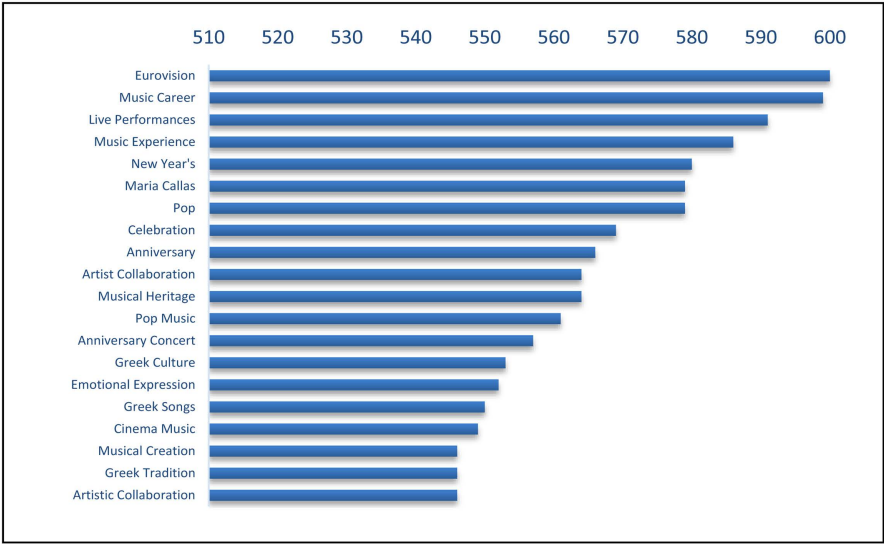


Figure 7. Bar chart of the method tags with the highest average outreach index.

4. Discussion

The tables and figures presented in the results section clearly show that the popularity of the classification tag itself and the popularity of the events that are described by such a tag in terms of outreach on the World Wide Web are significantly different and as such they present a challenge in terms of understanding the importance of each tag. This will be further discussed in this section.

4.1. RQ1. What Are the Most Popular Classifications of Audiovisual Art Events Regarding Event Type, Methods, and Subjects?

Classification tag popularity is a good indicator of how often an event that is described by that tag occurs in the audiovisual art event landscape. Tags that are popular represent types, methods and subjects of events that are commonplace, and that artists and event organizers are more interested in presenting to the audience. From the objective quantitative data collected in this study from the Web regarding contemporary audiovisual art events, it was made apparent that music dominated the event type classification tags, followed closely by the performing arts. In terms of techniques and methods musical events are represented by live singing and instrument performance, while the performing arts use acting,

dramaturgy and direction as their essential tools. In terms of themes, musical genres such as pop, rock, classical and laiko are popular, while well established story telling themes involving love, friendship and romance dominate in the performative arts.

Since the emergence of Web 2.0, music and musical events have played a predominant role in cultural dissemination on the Internet [28]. Participatory content diffusion combined with the popularity of musical events in younger and more dynamic audiences [29], has become a driving force behind the commonplace organization of such events and their multilayered Web presence that was in this research through real-world Web data. In an equivalent manner the performing arts are also utilizing Web 2.0 and attracting more dynamic audiences [30], thus establishing themselves a strong Web presence and an important place in the audiovisual art landscape.

From a musical standpoint genre popularity and the rise of pop music as a trend setter in the contemporary music landscape has been well established [31] [32] and is supported by the findings of this study as well. On the other hand, the worldwide importance of rap and hip hop [31] is not apparent in this research and is replaced instead by a slightly stronger representation of rock and an extraordinarily strong representation of Greek pop, folk and laiko music [27], thus highlighting the peculiarities of the contemporary Greek musical landscape. In the field of theater and cinema, the importance of themes closely related to personal relationships and feelings including love, romance, friendship as established by the popularity of the relevant tags in this research, is based on the importance of the emotional experience conveyed through the performing arts, which act as a link between the artist and the audience and their vital interdependency [33].

4.2. RQ2. What Are the Classifications of Audiovisual Art Events Regarding Event Type, Methods and Subjects That Are Most Influential Based on Event Outreach Index?

Going one step further this research focused not only on the estimated popularity of the various classification tags, but moreover on the outreach index of the events described by each such tag. With the use of objective real-world quantitative metrics derived from the Web the outreach index of each individual event was established and for each classification tag an average outreach index was calculated, indicating how impactful the various events that are described by that tag are. In the study of tat average outreach, general popular classification tags were replaced by specific tags that though their outreach indicate fewer but more influential events. In terms of event types, the importance of highly influential musical genres was made apparent, combined with an added importance to a distinct artistic identity. When it came to methods and techniques the high outreach of events that take advantage of technology in production was established, while the study of subject tags outlined the importance of the celebrational, annual and seasonal nature of audiovisual art events.

Globally important musical genres such as pop and rock displayed high influence as observed in other studies [31] [32], but were closely followed by genres closer to Greek idiosyncrasy as well less popular genres like classical and opera music, thus establishing how highly influential events of these genres are [34] despite being fewer. In terms of methods the high ranking of tags related to audio and visual effects and high production value was made apparent. Technologies and new media have a continuously significant importance in the performing arts worldwide [35] and are creating a new paradigm that attracts and impacts wide audiences. Finally, the study of subject tags showed a shift from commonplace themes to tags of a more temporal nature. Seasonal and annual events revolving around celebrations have been observed to be important social traditions that have important influence in local societies [36] and this is backed up by the findings of this study.

5. Conclusions

This study focuses on the use of contemporary real-world data, extracted from the Web through an advanced algorithm utilizing semantic web, search engine and generative AI technologies, for the purpose of classifying audiovisual art events. Using objective quantitative metrics, the event data were distilled into a singular web outreach index, which quantified their perceived impact based on their online presence. Using the events classification in terms of event type, methods and techniques, and thematic subjects, the popularity of various classification tags was explored. Going one step further, the impact of the events described by each tag was estimated using each event's outreach index. This process was used to discover important classification tags in terms of potential influence.

The analysis of the popularity of various classification tags within the current context of audiovisual art events in Greece, using a data-driven quantitative approach, objectively validated the significance of music and performative arts, alongside the popularity of particular musical genres and narrative themes focused on interpersonal relationships. Furthermore, examining the average outreach index for each classification tag highlighted the relevance of Greek musical heritage, the benefits of utilizing modern technology in performances, and the additional value associated with events that are annual, seasonal, or commemorative in nature. In addition to the insights discussed, the findings of the study encompass the methodology used, as well as the dataset collected, both of which can be applied similarly to yield data-driven discoveries to other research questions with different goals.

The focus of both the data extraction process and the classification and outreach analysis is the various events' online presence. The information available on the Web allows an objective numerical representation of the events but offers limited insight about events that may be important and culturally impactful, but disproportionately promoted online. This limitation of the study is somewhat mitigated by the extraction of information from user generated content, blogs, reviews etc.,

but remains important. For this reason, it is optimal to use data-driven quantitative analysis in tandem with qualitative expert analysis to combine findings and reach safer conclusions.

In the future, the collected dataset may be used in addressing issues beyond classification, looking into geographical or temporal attributes of audiovisual art events. Moreover, the analysis may move from a descriptive statistical discussion to the use of computational statistics and machine learning in order to gain deeper insight. In the next phase of this research, techniques such as clustering and topic modeling will be used to gain better understanding of the dataset, based both on its quantitative characteristics and its descriptive texts.

Obtaining a clearer understanding of the landscape of audiovisual art events using real-world quantitative metrics, in tandem with a critical qualitative approach of the wider field of art and culture, can provide the basis for more informed decision making from, the artists, the organizers and the audience. This clearer approach may act for the benefit of everyone involved, leading to more appealing, interesting, and lucrative events and to a deeper relationship between the artistic product and the general public.

Funding

This research was funded by the Hellenic Foundation for Research & Innovation, within the framework of the “2nd Call for H.F.R.I.’s Research Projects to Support Faculty Members & Researchers” (project number 3607).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H.F. and Secret, A. (2023) The World-Wide Web. In: Seneviratne, O. and Hendler, J., Eds., *Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web*, ACM, 51-65. <https://doi.org/10.1145/3591366.3591373>
- [2] Jensen, K.B. and Craig, R.T. (2016) *The International Encyclopedia of Communication Theory and Philosophy*. Wiley.
- [3] Pergantis, M., Varlamis, I., Kanellopoulos, N.G. and Giannakouloupoulos, A. (2023) Searching Online for Art and Culture: User Behavior Analysis. *Future Internet*, **15**, Article 211. <https://doi.org/10.3390/fi15060211>
- [4] Barifah, M., Landoni, M. and Eddakroui, A. (2020) Evaluating the User Experience in a Digital Library. *Proceedings of the Association for Information Science and Technology*, **57**, e280. <https://doi.org/10.1002/pra2.280>
- [5] Gaona-García, P.A., Martin-Moncunill, D. and Montenegro-Marin, C.E. (2017) Trends and Challenges of Visual Search Interfaces in Digital Libraries and Repositories. *The Electronic Library*, **35**, 69-98. <https://doi.org/10.1108/el-03-2015-0046>
- [6] Leibovitz, T., Roig, A. and Sánchez-Navarro, J. (2015) Up Close and Personal: Exploring the Bonds between Promoters and Backers in Audiovisual Crowdfunded Projects. In Bennett, L., Chin, B. and Jones, B., Eds., *Crowdfunding the Future Media*

- Industries, Ethics, and Digital Society*, Peter Lang, 15-30.
- [7] Maier, S. (2010) All the News Fit to Post? Comparing News Content on the Web to Newspapers, Television, and Radio. *Journalism & Mass Communication Quarterly*, **87**, 548-562. <https://doi.org/10.1177/107769901008700307>
 - [8] Boczkowski, P.J. and Mitchelstein, E. (2013) The News Gap: When the Information Preferences of the Media and the Public Diverge. The MIT Press. <https://doi.org/10.7551/mitpress/9780262019835.001.0001>
 - [9] Martens, B., Aguiar, L., GGmez, E. and Mueller-Langer, F. (2018) The Digital Transformation of News Media and the Rise of Disinformation and Fake News. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3164170>
 - [10] Nadotti, L. and Vannoni, V. (2019) Cultural and Event Tourism: An Interpretative Key for Impact Assessment. *Eastern Journal of European Studies*, **10**, 115-131.
 - [11] Giannakouloupoulos, A., Pergantis, M., Konstantinou, N., Kouretsis, A., Lamprogeorgos, A. and Varlamis, I. (2022) Estimation on the Importance of Semantic Web Integration for Art and Culture Related Online Media Outlets. *Future Internet*, **14**, Article 36. <https://doi.org/10.3390/fi14020036>
 - [12] Jacquette, D. (2014) *Ontology*. Routledge.
 - [13] O'Leary, D. (2005) Review: Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. *The Computer Journal*, **48**, 498-498. <https://doi.org/10.1093/comjnl/bxh100>
 - [14] Berners-Lee, T., Hendler, J. and Lassila, O. (2001) The Semantic Web. *Scientific American*, **284**, 34-43. <https://doi.org/10.1038/scientificamerican0501-34>
 - [15] Berners-Lee, T., Hall, W., Hendler, J.A., O'Hara, K., Shadbolt, N. and Weitzner, D.J. (2006) A Framework for Web Science. *Foundations and Trends® in Web Science*, **1**, 1-130. <https://doi.org/10.1561/18000000001>
 - [16] Gandon, F. and Schreiber, G. (n.d.) RDF 1.1 XML Syntax. <https://www.w3.org/TR/rdf-syntax-grammar/>
 - [17] Beckett, D., Carothers, G. and Seaborne, A. (2014) RDF 1.1 N-Triples: A Line-Based Syntax for an RDF Graph. W3C Recommendation.
 - [18] Beckett, D., Berners-Lee, T., Prud'hommeaux, E. and Carothers, G. (2014) RDF 1.1 Turtle. World Wide Web Consortium, 18-31.
 - [19] Sporny, M., Longley, D., Kellogg, G., Lanthaler, M. and Lindström, N. (2020) JSON-LD 1.1. W3C Recommendation.
 - [20] Maybury, M.T. (1995) Generating Summaries from Event Data. *Information Processing & Management*, **31**, 735-751. [https://doi.org/10.1016/0306-4573\(95\)00025-c](https://doi.org/10.1016/0306-4573(95)00025-c)
 - [21] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., et al. (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
 - [22] Bhosale, H.S. and Gadekar, D.P. (2014) A Review Paper on Big Data and Hadoop. *International Journal of Scientific and Research Publications*, **4**, 1-7.
 - [23] von Bloh, J., Broekel, T., Özgün, B. and Sternberg, R. (2019) New(s) Data for Entrepreneurship Research? An Innovative Approach to Use Big Data on Media Coverage. *Small Business Economics*, **55**, 673-694. <https://doi.org/10.1007/s11187-019-00209-x>
 - [24] Rosenbush, S. and Totty, M. (2013) How Big Data Is Changing the Whole Equation for Business. *Wall Street Journal*.
 - [25] Manovich, L. (2016) The Science of Culture? Social Computing, Digital Humanities

- and Cultural Analytics. *Journal of Cultural Analytics*, **1**, 1-15.
<https://doi.org/10.22148/16.004>
- [26] Giannakouloupoulos, A., Pergantis, M., Lamprogeorgos, A., Lampoura, S., Limniati, L., (2024) Presenting Web-Extracted Contemporary Audiovisual Arts Events with a Focus on User Experience (UX). EUTIC 2024 (Under Publication).
 - [27] Economou, L. (2018) Sentiment, Memory, and Identity in Greek Laiko Music (1945-1967) In: Tragaki, D., Ed., *Made in Greece*, Routledge, 17-27.
<https://doi.org/10.4324/9781315749075-3>
 - [28] Beer, D. (2008) Making Friends with Jarvis Cocker: Music Culture in the Context of Web 2.0. *Cultural Sociology*, **2**, 222-241. <https://doi.org/10.1177/1749975508091034>
 - [29] Robinson, R. (2016) Music Festivals and the Politics of Participation. Routledge.
<https://doi.org/10.4324/9781315596785>
 - [30] Coleman, L.J., Jain, A., Bahnan, N. and Chene, D. (2019) Marketing the Performing Arts: Efficacy of Web 2.0 Social Networks. *Journal of Marketing Development and Competitiveness*, **13**, 23-28.
 - [31] Petitbon, A.M. and Hitchcock, D.B. (2022) What Kind of Music Do You Like? A Statistical Analysis of Music Genre Popularity over Time. *Journal of Data Science*, **20**, 168-187. <https://doi.org/10.6339/22-jds1040>
 - [32] Fossi, J., Dzwonkowski, A. and Othman, S. (2021) Analyzing Music Genre Popularity. In: Arai, K., Ed., *Intelligent Computing*, Springer International Publishing, 284-294.
https://doi.org/10.1007/978-3-030-80126-7_22
 - [33] Soltani, M. (2014) The Essence of Eddie: A Reflection of the Interplay between Art, Relationships, and Emotion. *Academic Psychiatry*, **38**, 751-751.
<https://doi.org/10.1007/s40596-014-0227-3>
 - [34] Getz, E. (2015) *Why Classical Music Still Matters*. University of California Press.
 - [35] Hebert, D., Kallio, A.A. and Odendaal, A. (2012) Not *So* Silent Night: Tradition, Transformation and Cultural Understandings of Christmas Music Events in Helsinki, Finland. *Ethnomusicology Forum*, **21**, 402-423.
<https://doi.org/10.1080/17411912.2012.721525>
 - [36] Todi, C. (2019) The Metamorphosis of Performing Arts. *Theatrical Colloquia*, **9**, 173-186. <https://doi.org/10.2478/tco-2019-0004>