SPECIAL ARTIFICIAL INTELLIGENCE ISSUE

JUNE 2022

Trends and Issues in Library Technology

International Federation of Library Associations

IFLA IT Section

TRENDS & ISSUES IN LIBRARY TECHNOLOGY

IFLA Information Technology Section — Special Issue on Artificial Intelligence

Feature Articles

IFLA Summer Dublin Conference

Departments

Editor's Notes

Special Artificial Intelligence Issue

Dear Colleagues,

I am very happy to introduce this special AI themed *Trends and Issues in Library Technology* with a great cross-section of AI focused articles and important IFLA IT section conference information. For the conference, there is information on IFLA IT-sponsored sessions, co-sponsored IT panels with other sections, SIGS and satellite conferences also co-sponsored with the group. In keeping with this issue's AI theme, I want to highlight an exciting AI satellite conference preceding the WLIC which will take place in Galway, Ireland and encourage you to attend. See https://ucg.ie/ifla/ and the Satellite program in this issue (p. 37) for more information.

IFLA WLIC 2022 DUBLIN

Inspire, Engage, Enable, Connect

87th World Library and Information Congress
26-29 July 2022, Dublin, Ireland

For this issue, we have a wide spectrum of great articles on AI topics for libraries. Juja Chakarova, of the International Atomic Energy Agency, Vienna, gives a historical overview of AI in the context of libraries. Lynn Kleinveldt, Cape Peninsula University of Technology, South Africa, overviews new AI learning environments and libraries. Elena Sánchez Nogales, Alicia Pastrana García, José Carlos Cerdán Medina, of the National Library of Spain, present a project regarding Digital Transformation, Data Reuse and AI at the National Library of Spain. Marcelo Lorca González, from the Library at the National Congress of Chile, details a project on the conversion of MARC to BIBFRAME and implications towards semantic web-linking. From Greece's International Hellenic and Aristotle University of Thessaloniki and Datascouting, Afrodite Malliari, Ilias Nitsos, Sofia Zapounidou, Stavros Doropoulos overview an AI-based Open Audio-Visual Archives (OAVA) project for enhanced search and retrieval. And from the University of Sheffield, UK, Andrew Cox forwards a compelling set of ethics scenarios regarding artificial intelligence and the importance of examining ethical implications of AI for information professionals.

Our chair, Edmund Balnaves, also gives a quick update on the IT section's 2022 progress and WLIC 2022 conference information. To better kick-off our AI theme, I have written an editorial on our AI articles, drawing together connections and introducing themes and authors.

I would also like to thank our wider Communications team: first, IT Section Communications Director, Francois-Xavier Boffy for help with layout; and Wouter Klapwijk, South Africa, for his valuable editing skills. Finally, a special note of appreciation to all our section authors. They have all done amazing work contributing articles to this special AI focused issue. I would also mention that if you are looking to volunteer or help with TILT, we are searching for a skilled volunteer graphic designer/layout artist who can speed up our publication processes. We are also always looking for upcoming high caliber contributions and projects. Our submissions door is now open again, so feel free to submit a proposal or query for upcoming issues. Also, if you have not joined online already, I encourage all of you to visit, join and contribute to our new IFLA IT Social Media Facebook Group available at https://www.facebook.com/groups/iflaitsection

I hope everyone will be attending the upcoming conference in Dublin. I had recently read that the top five things to do and see in Dublin are: 1) Trinity College, one of the oldest universities in the world; 2) Grafton Street where you can shop till you drop; 3) St. Stephen's Green, Dublin's Central park where writers James Joyce and W.B. Yeats have memorials commemorating contributions to Irish poetry and literature; 4) The National Gallery of Ireland; 5) Guinness Brewery and Storehouse.

To all our readers, I look forward to meeting you at our upcoming annual conference in Dublin and I extend my best wishes for safe travels.

Sincerely,

Ray

Ray Uzwyshyn, Ph.D. MLIS ruzwyshyn@txstate.edu
Editor, Trends and Issues in Library Technology

The Development of an audio-visual aggregator: the Open Audio-Visual Archives (OAVA) project in Greece

Afrodite Malliari, afmalliari@gmail.com
International Hellenic University, Department of Library, Archival & Information Studies

Ilias Nitsos, initsos@gmail.com
International Hellenic University, Department of Library, Archival & Information Studies

Sofia Zapounidou, zapounidous@gmail.com
Aristotle University of Thessaloniki, Library & Information Centre

Stavros Doropoulos, doro@datascouting.com
DataScouting

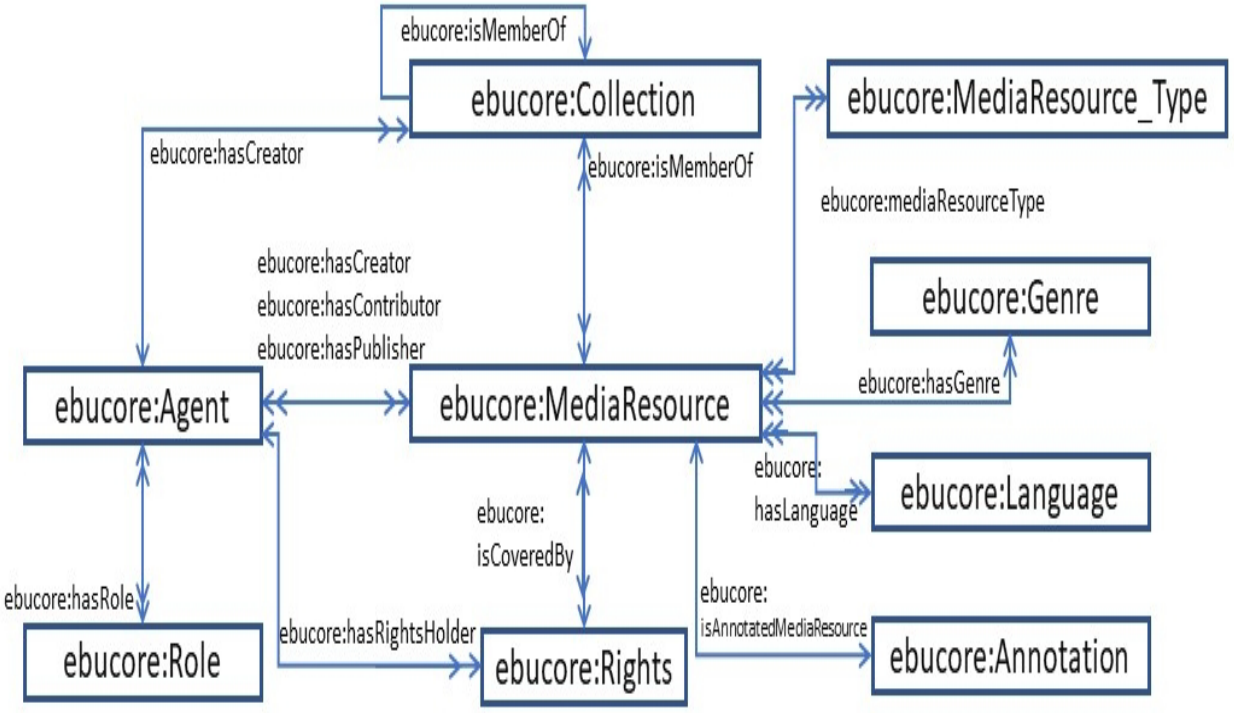

Figure 1. The OAVA Model depicting classes, and the relationships between them

Abstract

The available online audiovisual material maintained by public and private organizations and institutions is numerous and ranges from videos and narrations regarding historical and everyday life events to scientific, academic, and cultural events. Despite the apparent availability of audiovisual resources, finding useful resources is not as easy as one might think. Poor indexing by search engines and lack of metadata schemes are the main reasons for this inconvenience. The problem of dispersion of audiovisual resources and diversity of resource providers is largely solved by aggregation services.
In Greece there is no reference point for the search and access to audiovisual material. A Greek National Registry of audiovisual providers has not yet been implemented. The Open Audio-Visual Archives (OAVA) project aims to gather audiovisual material that is of Greek interest or contains Greek speech. The OAVA project provides a unified search mechanism not only to the aggregate metadata of audio-visual material but also to its searchable content. Through the application of deep learning models, algorithms are developed that perform Automatic Speech Recognition in Greek and in English.

During the project, 500 Greek-language audiovisual content providers were reviewed. 233 of them were found eligible according to specific criteria. The metadata used by each provider were mainly application schemes without metadata schemes. Those were recorded and mapped with Vufind and with EBUCore schema. The open-source software VuFind was configured to operate as the basis of the unified search platform.

This article briefly presents the objectives of the project, selection criteria, results for content and audiovisual providers in Greece. It overviews licensing issues, the OAVA metadata scheme and the basic functions of the search mechanism.

KEYWORDS

Audiovisual material, speech to text technologies, cultural heritage, open access, content aggregators

Introduction

The plethora of web content nowadays, including audiovisual content, is constantly creating challenges in terms of harvesting, indexing and searching. In the case of audiovisual content, much of the existing material remains unused, because it cannot be detected by potential visitors. In order to provide a single point of access, the creation of aggregation services is the only option for national or international organizations that wish to promote and make better use of online audiovisual resources. A typical example is Europeana (c2022), that focuses on providing European cultural heritage institutions with the necessary tools that will enable them to effectively share their collections (Europeana Foundation, 2019). Another effort is Trove (c2022), that aggregates content from "libraries, museums, galleries, the media, government and community organizations" in Australia (trove, c2022). Memobase (Memoriav, c2022) and DigitalNZ (National Library of Zealand, c2022) with similar examples in Switzerland.

The article examines the need for the development of an aggregator for content offered by Greek providers. It describes the steps to create this type of aggregator, as part of the OAVA project. Deep Learning models are used to create searchable text of the audiovisual content in Greek and English. An aggregator not only collects metadata, but also provides speech search mechanisms.

Research

The first step in creating the Open Audio-Visual Archives (OAVA) platform to provide access to various audiovisual resources was to create a multidisciplinary dataset. The dataset should contain resources, free of copyright limitations, with informative content excluding literary or artistic. The OAVA project aims to provide access not only to aggregated metadata but also to searchable content. It apples deep learning models and converts the speech contained in the audiovisual material into searchable text.

As there is no national searchable registry of audiovisual material in Greece, we gathered resources by:

a) conducting an online survey to Greek reference librarians who in some cases suggested and provided audiovisual resources to their users

b) searching with keywords on the web (e.g. "local history" AND "video")

c) browsing the websites of organizations that were expected to provide audiovisual content due to their activities and functions (Malliari, et.al., 2022).

A study of mapping collections of audiovisual resources in Europe (Klijn, & de Lusenet, 2008) was considered in order to select the websites.

The selection process took place in the first quarter of 2021. 500 providers of Greek-language audiovisual content were evaluated using the crAAp test (Currency, Relevance, Authority, Accuracy, and Purpose). If a website was eligible for further evaluation, its resources were evaluated with specific selection criteria based on:

(a) the Digital Library Federation and Council on Library and Information Resources study (Pitschmann, 2001) and (b) criteria found in other studies (Digital Public Library of America, c2022; Oesterlen, 2017; Scholz, 2019; Trove, c2020; Turnok, Kaye, & Carrasqueiro, 2010).

Criteria were grouped under context, content, form/use, process or technical, and metadata.

A total of 497 of the 500 candidate providers evaluated were found to be reliable by applying the crAAp test to their content. The review process was completed with the specific selection criteria mentioned. Eventually 233 providers were found eligible and included in the final trusted list containing: libraries, archives, museums, universities, governmental and non-governmental organizations and media organizations. More than half of them (58%) came from the public sector and universities. Universities contribute most of the content.

Approximately 80% of the total of 1710 resources offered came from universities. It is also interesting to mention that the majority of the resources were "open courses" (77%). Other resource categories were: "open educational material", "academic and/ or scientific events", "cultural events", "Interviews", "board meetings", "Campaigns", "archival material".

In terms of licensing, Greek audiovisual content providers tend to choose the same licensing method for all their resources. More than half of them chose to publish their resources in platforms. Their content can be reused under the terms of "Fair Use" (66%). 28% of the providers prefer to publish under "Creative Commons" licenses. 5% of providers have their own terms of use that do not match the existing licensing typology. Another small percentage (1%) use rights statements.

The Malliari, et.al. research (2022) revealed that most of the providers do not own any kind of publishing infrastructure. They utilize commercial streaming services, such as YouTube. In terms of metadata policy, almost every eligible provider (97%) uses an application profile developed according to their own needs and/or the capabilities of the information system they use. They do not utilize a well-known metadata schema, like DC. The dataset with the above-mentioned information regarding provider's name and type, channel publishing AV, resource genre and URL, licensing and metadata can be found at Zenodo[1].

OAVA schema

The OAVA project aims to collect online audiovisual content. Technical collaboration with stakeholders is not guaranteed. This is unlike other aggregation services such as Europeana and Digital Public Library of America . To aggregate and display the resources metadata into the OAVA platform, the metadata must conform to the OAVA schema.

[1] https://doi.org/10.5281/zenodo.5112283

The OAVA schema is based on the EBUCore Schema. The EBUCore Schema has been developed by the European Broadcasting Union for the description of broadcasting resources (EBU, 2020). It conforms to the EBU Class Conceptual Data Model (CCDM). It also takes other models into account (e.g., FRBR, Europeana Data Model) to enable mappings. The OAVA Schema uses only a subset of the EBUCore Schema. The descriptive needs are different in the OAVA project compared to the broadcasting resources on which the EBUCore Schema focuses.

The OAVA Schema uses 4 main classes: Media Resource, Collection, Rights, and Agent. The Media Resource class is used to describe the audiovisual resource providing information about the title, language, URL, abstract, thumbnail, etc. In case the Media Resource is part of a collection, the Collection class is used to provide the title and the description of the collection. The Rights class provides information about the rights by which a Media Resource instance is covered. The Agent class describes the agent. This is either the corporate body or person that has created, contributed to, or published the Media Resource. The OAVA Schema also uses other classes that use controlled vocabularies (i.e., Media Resource Type, Genre, Language). The remaining Annotation class is used to describe the enrichment of the metadata by the OAVA project. In the following figure, the OAVA model is displayed in Figure 1 above. Classes are depicted as rectangles.

The relationships between classes are depicted as arrows. The type of the relationship (1-1, 1-M, M-N) and cardinalities are also presented. To instantiate the model, metadata from providers are aggregated in a VUFind instance. They are then exported using the OAVA Schema conceptualizations. The metadata from providers are aggregated from publishers' webpages or systems. This often uses heuristic methods due to the lack of metadata standards. The aggregated metadata are enriched and mapped to the VUFind Schema that has been extended to conform to the OAVA Schema (see figure 2). Some examples of metadata enrichment involve the creation of a thumbnail for the resource, the assignment of type (e.g., video or audio), the provider's logo, the language of the resource, etc. The VUFind Schema consists of more than 100 elements, of which some are repeatable. For the needs of the OAVA project, these elements were analyzed to discover correspondences to the elements of the OAVA schema. When such correspondences were not found, the needed OAVA Schema elements (subset of the EBUCore Schema) were added to the VUFind schema. For the needs of the OAVA project, 22 elements from the VUFind were selected as corresponding to OAVA schema properties. 16 properties from the OAVA Schema were added (all under the ebucore namespace).

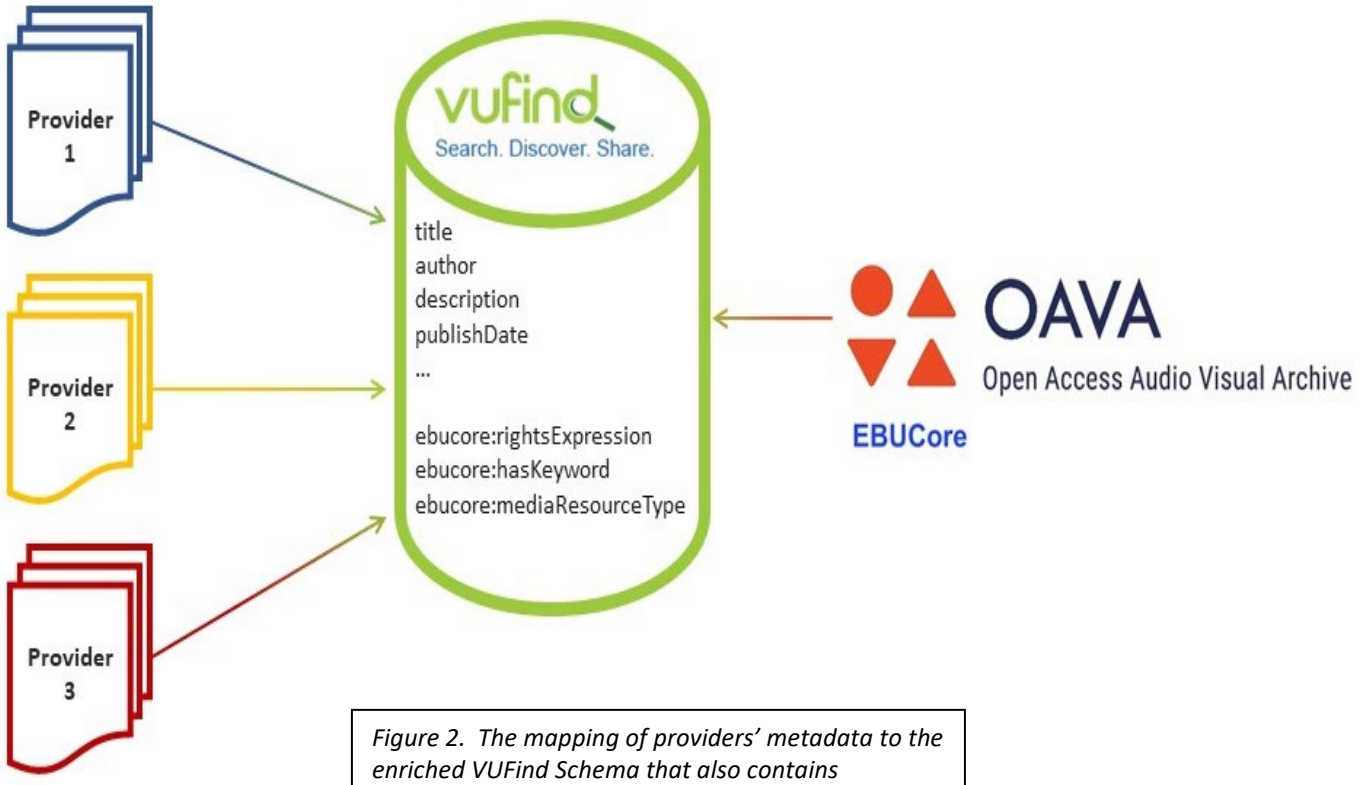

Figure 2. The mapping of providers' metadata to the enriched VUFind Schema that also contains properties from the OAVA Schema.

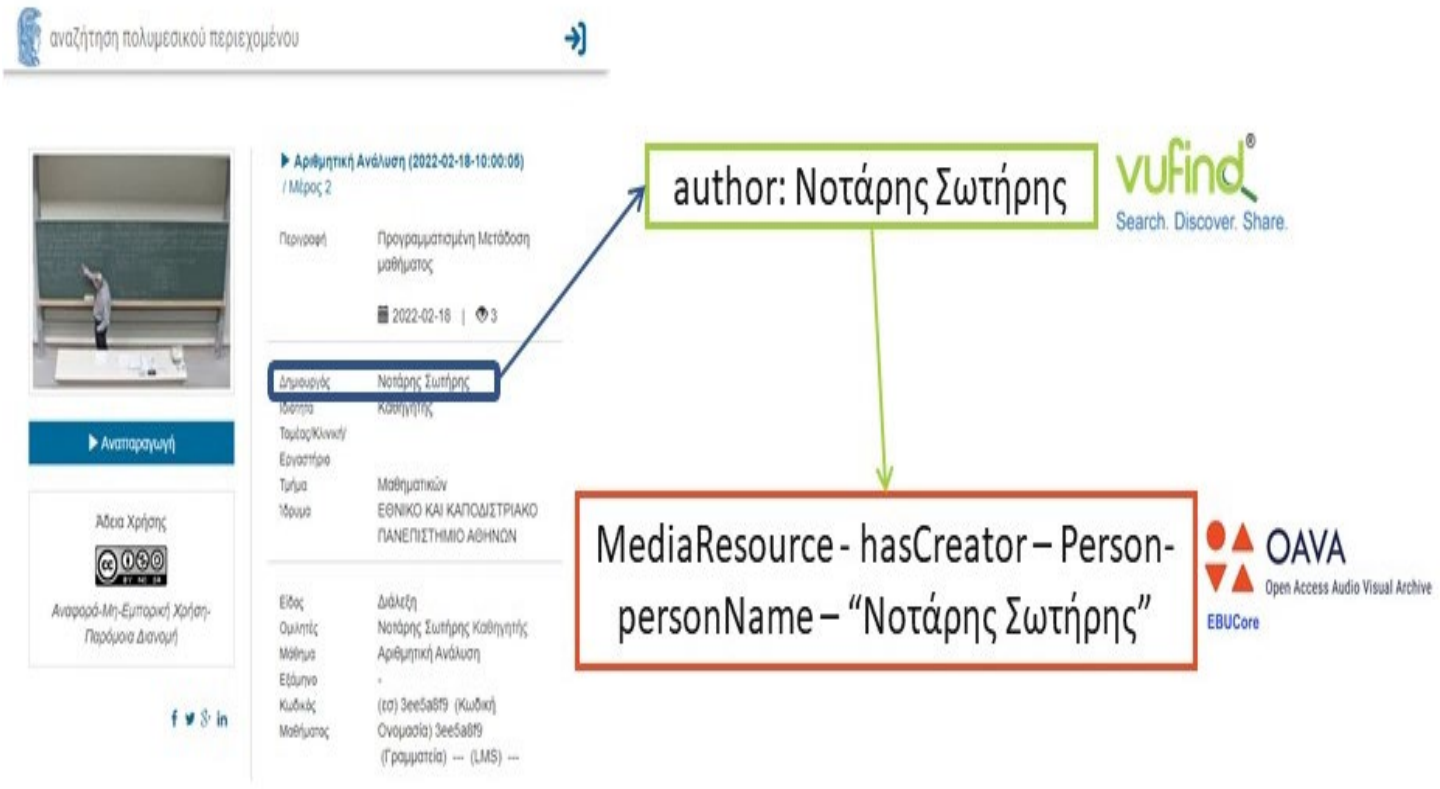

Figure 3. Author information from the provider's site is mapped to the VUFind element author. This piece of information triggers the creation of two triples MediaResource - hasCreator - Person and Person - personName - 'Νοτάρης Σωτήρης'

.

After the metadata is aggregated and stored in the VUFind Schema, it can be exported as triples using the OAVA Schema. Each record in VUFind describes a *MediaResource*. From the value of one property more than one triple can be created. In the following example, a lecture is described having as an author the *Person* 'Νοτάρης Σωτήρης'. After mapping to the OAVA Schema and exporting in triples, two triples are created: *MediaResource - hasCreator - Person* and *Person - personName* - 'Νοτάρης Σωτήρης' (Figure 3).

OAVA platform

A general overview of the OAVA platform architecture is presented below. Basic components and processing stages of the OAVA platform are now discussed. (Detailed information is included for the Automatic Speech Recognition approach, especially for Greek content).
The orchestrator is a basic component of the OAVA platform. It is used to monitor sources of audiovisual content. When the orchestrator finds new content that is missing from OAVA, it initiates a crawling procedure. The Worker is another basic component of the platform. It can crawl pages using either a traditional or headless browser that does not have a graphical user interface. While the crawling process is going on, the crawled pages are sent to a scraping procedure. A messaging agent, RabbitMQ, has been used to communicate between OAVA services. This allows OAVA to be horizontally scalable and distributed in different systems.
The scraping procedure plays a significant role in the functionality of the platform. The scraping determines whether the crawled page contains audiovisual material that needs to be indexed and analyzed. Once it finds such material it begins to extract useful information (title, description, authors, publishers, video links, etc.). The OAVA scraper maps each different field into a common model. Eventually, all the scraped information originating from different sources is mapped to the same format.
After the completion of the scraping procedure, the produced scraped information is sent back to the orchestrator. It is prepared for publishing and further enriched with metadata. The orchestrator then finds any multimedia links attached to this scraped information. It then forwards these links to a storage manager which the storage manager evaluates. The storage tries to find the best approach to get the multimedia content attached to that link. As soon as the downloading process of these multimedia files has been completed, the storage manager sends back the locally stored file paths to the orchestrator. The orchestrator then publishes the information containing the multimedia file paths to the main platform.

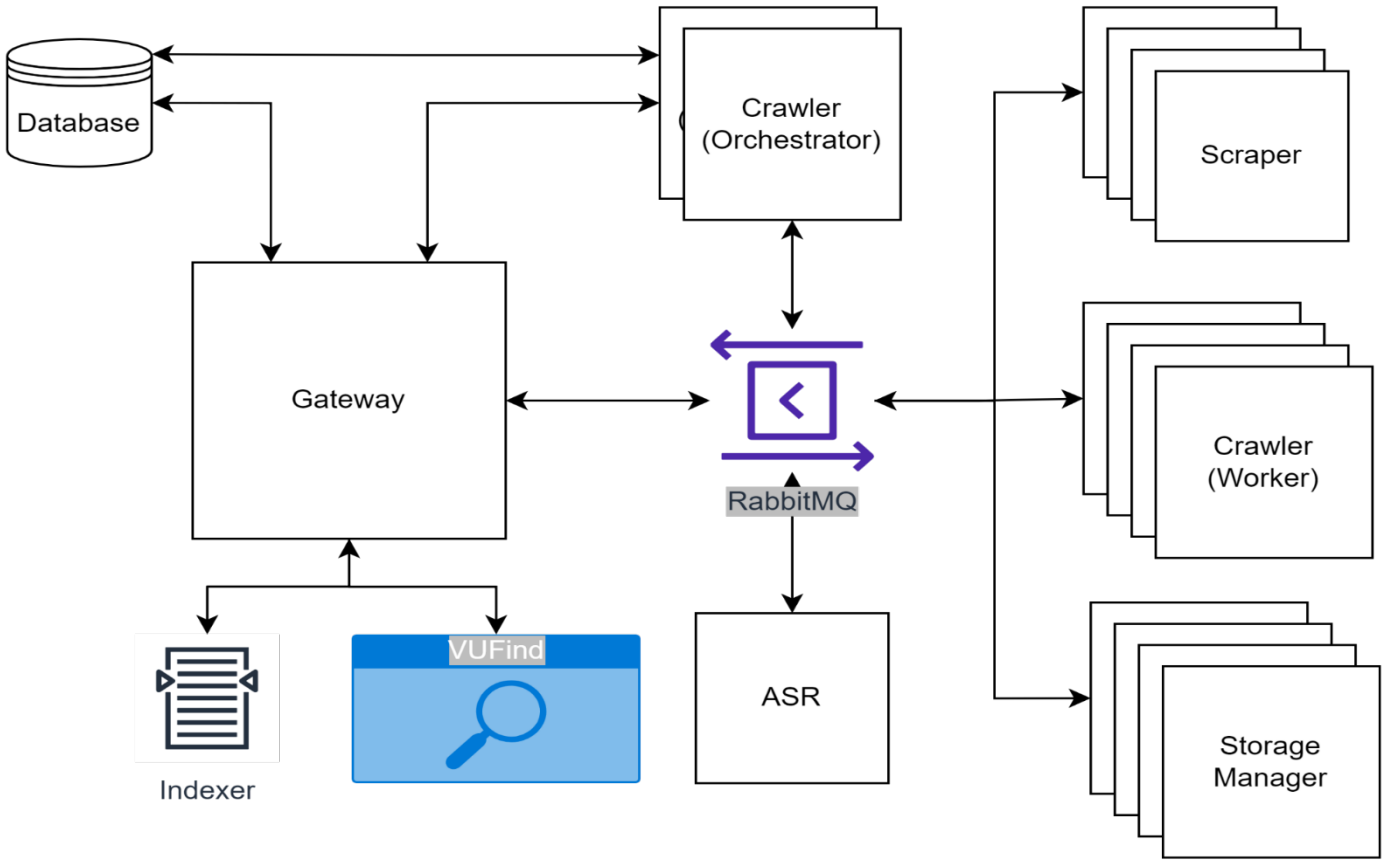

Figure 4. Overview of the OAVA platform architecture

Automatic Speech Recognition

One of the main jobs of the OAVA platform is to extract text. This is from the downloaded multimedia files using Machine Learning and more specifically Automatic Speech Recognition techniques. Automatic Speech Recognition models were developed for both English and Greek content. Models including datasets with over 7000 hours of oral speech in total were used Especially for the English language, (e.g., Librispeech, Fisher Corpus, Switchboard-1 Dataset, WSJ-0 and WSJ-1, National Speech Corpus - 1, Mozilla Common Voice). The top performing models also required a lot of resources for inference. For the English language we selected architectures that require less resources without significantly sacrificing performance. Examples of such models are CTC architectures (Graves et al., 2006) like Jasper (Li et al., 2019) and QuartzNet (Kriman et al., 2019). As a post-processing step we use a 4-gram language model to improve the output of ASR models.

For the Greek language there are no big datasets available. Training such models requires large scale speech recognition datasets. These consist of speech in audio format and the corresponding transcription. A few examples of Greek datasets are: Greek Mozilla Common Voice (22 hours), Transcribedbook (4 hours), Greek TEDx speeches (20 hours). It is apparent that the size of this corpus is very small compared to English datasets. To this end, transcribing additional speech data is required for training a Greek ASR model. In OAVA, more than 100 hours of additional data was transcribed and made available for training. This dataset includes: audio from the Greek parliament channel, Greek news channels, university open lectures, Onassis foundation events, NGO events and library events. After combining these datasets training of various state-of-the-art architectures took place for testing purposes. Note that for each experiment the weights of the English pretrained model were used as a starting point as preliminary experiments revealed better performance in the final Greek model when starting the training from an English pretrained model, instead of random weight initialization. For the evaluation dataset, the audio from the university open lectures was merged with Onassis foundation event, NGO events and library events. The focus was mainly on two architecture categories, namely the Connectionist Temporal Classification (CTC) models (Graves et al., 2006) and RNN-Transducers (RNN-T) (Graves et al., 2012). For CTC models, apart from Jasper and QuartzNet experiments were also conducted with Citrinet (Majumdar et al., 2021). For the RNN-T models experiments were conducted with ContextNet (Han et al., 2020) and Conformer (Gulati et al., 2020).

The overall WER of the transducer model was further improved by adding a language model as a post processing step. To do this a 6-gram language model was trained using millions of articles that were crawled from Greek blogs. A Greek language model was trained that was applied after the conformer transcription. The final WER was close to 0.20 in the evaluation set.

When audio arrives in the Automatic speech recognition module it is processed. The corresponding English or Greek model is then applied to produce transcription. As soon as the process is finished the new enhanced information enriches the existing one. The multimedia files used in the process are then permanently deleted. After all this process, the produced information is exported and mapped to VUFind.

Conclusions

This article has given a high-level outline of a research project aimed at developing a platform that will aggregate audiovisual resources by Greek providers. As search results give access not only to aggregated metadata (using the OAVA model), but also searchable audio content, resources should be free of copyright limitations. In this respect, only informative content has been selected (not literary or artistic). Deep learning models have been applied that convert the speech contained in the audiovisual material into searchable text to enhance the user search experience.

Funding

This work was supported by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH - CREATE - INNOVATE [project code: T2EDK-00526].
Project partners: Department of Library, Archival and Information Studies, International Hellenic University; DataScouting, Software Development Company

References

Digital Public Library of America. (c2022). *Collection Development Guidelines*. Available from: https://pro.dp.la/hubs/collection-development-guidelines
EBU. (2020). *EBU Core Metadata Set (EBUCore): Specification v. 1.10. TECH 3293*. Available from: https://tech.ebu.ch/docs/tech/tech3293.pdf
Europeana. (c2022). *Europeana.* Available from: https://www.europeana.eu/
Europeana Foundation. (2019). *Europeana Publishing Framework 2.0.* Available from:https://pro.europeana.eu/files/Europeana_Professional/Publications/Publishing_Framework/Europeana%20Publishing%20Framework%20V2.0%20English.pdf
Graves, A. (2012). *Sequence transduction with recurrent neural networks*. arXiv preprint arXiv:1211.3711.
Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006, June). Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning* (pp. 369-376). doi: 10.1145/1143844.1143891
Gulati, A., Qin, J., Chiu, C. C., Parmar, N., Zhang, Y., Yu, J., ... & Pang, R. (2020). *Conformer: Convolution-augmented transformer for speech recognition*. arXiv preprint arXiv:2005.08100.
Han, W., Zhang, Z., Zhang, Y., Yu, J., Chiu, C. C., Qin, J., ... & Wu, Y. (2020). *Contextnet: Improving convolutional neural networks for automatic speech recognition with global context*. arXiv preprint arXiv:2005.03191.
Klijn, E., & de Lusenet, Y. (2008). *Tracking the reel world: A survey of audiovisual collections in Europe*. Amsterdam: European Commission on Preservation and Access. Available from: https://www.ica.org/sites/default/files/WG_2008_PAAG-tracking_the_reel_world_EN.pdf
Kriman, S., Beliaev, S., Ginsburg, B., Huang, J., Kuchaiev, O., Lavrukhin, V., ... & Zhang, Y. (2020, May). Quartznet: Deep automatic speech recognition with 1d time-channel separable convolutions. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6124-6128). IEEE. doi: 10.1109/ICASSP40776.2020.9053889
Li, J., Lavrukhin, V., Ginsburg, B., Leary, R., Kuchaiev, O., Cohen, J. M., ... & Gadde, R. T. (2019). *Jasper: An end-to-end convolutional neural acoustic model*. arXiv preprint arXiv:1904.03288.
Majumdar, S., Balam, J., Hrinchuk, O., Lavrukhin, V., Noroozi, V., & Ginsburg, B. (2021). *Citrinet: Closing the Gap between Non-Autoregressive and Autoregressive End-to-End Models for Automatic Speech Recognition*. arXiv preprint arXiv:2104.01721.
Malliari, A., Nitsos, I., Zapounidou, S., & Doropoulos, S. (2022). Mapping audiovisual content providers and resources in Greece. *International Journal on Digital Libraries*, 1-11. doi: 10.1007/s00799-022-00321-6
Memoriav. (c2022). *Memobase: Access to Switzerland's audiovisual cultural heritage*. Available from: https://memoriav.ch/en/memobase
National Library of New Zealand. (c2022). *About DigitalNZ.* Available from: https://digitalnz.org/about
Oesterlen, E.-M. (2017). *Aggregation Handbook, 3rd edition*. The Hague: Europeana, EUscreen. Available from: http://blog.euscreen.eu/wp-content/uploads/2017/03/Aggregation-Handbook_Revised-edition3_2017.pdf
Pitschmann, L.A. (2001). *Building Sustainable Collections of Free Third-Party Web Resources*. Washington, D.C.: Digital Library Federation, Council on Library and Information Resources. Available from: https://clir.wordpress.clir.org/wp-content/uploads/sites/6/pub98_57d70f70b208f.pdf
Scholz, Henning. (2019). *Europeana Publishing Guide v1.8: A guide to the metadata and content requirements for data partners publishing material in Europeana Collections*. The Hague: Europeana Foundation. Available from: https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Publishing%20Guide%20v1.8.pdf
Trove. (c2020). *Technical specifications*. Available from: https://trove.nla.gov.au/technical-specifications
Trove. (c2022). *What is Trove.* Available from: https://trove.nla.gov.au/about/what-trove
Turnok, R., Kaye, L., & Carrasqueiro, L. (2010). EUscreen content selection policy. In *EUscreen Workshop, 23-24 June, Mykonos, Greece*. Available from: https://www.slideshare.net/EUscreen/turnok-kayecarrasqueiro-e-uscreen-content-selection-policy-euscreen-mykonos