# Multimedia Content Annotations for Rapid Exploitation in Multi-Screen Environments: The MECANEX project vision

Eirini Eleni Tsiropoulou, Eleni Stai, Stella Kafetzoglou, Konstantinos Rapantzikos, Symeon Papavassiliou

Institute of Communication and Computer Systems (ICCS)
Athens, Greece

Miggi Zwicklbauer

FOKUS Gesellschaft Zur Foerderung
Der Angewandten Forschung E.V,
Germany

Pieter van Leeuwen

Noterik B. V.
Amsterdam, Netherlands

Vassiliki Karampali

Motivian
Athens, Greece

***Abstract*— As the prerequisites of production houses, broadcasters, advertising agencies and online publishing companies for enriched multimedia content increase rapidly, the need of innovative tools for the creation of enriched multimedia content is undeniable. Indeed, the adaption of enriched multimedia content to multi-screen environments, which enables the automatic porting to different target platforms (e.g. regular web pages and TV applications), is a significant asset for the global market of enriched multimedia content. Within this framework, in this paper we propose the Multimedia Content Annotations for Rapid Exploitation in Multi-Screen Environments (MECANEX) toolkit that consists of the following key components/tools: a) the automatic annotation and editorial tool, b) the multi-screen tool and c) the social and personalization tool. The MECANEX toolkit offers the ability to automatically annotate, edit and semi-automatically enrich video collections for multi-screen applications, as well as rapidly create and synthesize multimedia content, prototype multi-screen concepts and test related interaction models. Personalization and recommendation schemes are leveraged towards improving end-users' experience via creating and delivering to them enriched multimedia content aligned to their preferences and needs. The MECANEX toolkit's upper aim is to facilitate and allow the development of solid business models, marketing and advertising campaigns targeted to users' needs and preferences.**

***Keywords—* automatic annotation; multi-screen; personalization; relevance feedback; recommendation; editorial tool; enrichments; advertisements***

## I. INTRODUCTION

With the growing demand of production houses, broadcasters, advertising agencies and online publishing companies for enriched multimedia content, the development of innovative tools supporting automatic annotation, editorial processes, personalized functionalities and synchronization in multi-screen environments becomes a need rather than a desire. The wide range of available multimedia devices and access technologies, along with the ubiquity of mobile portable devices offering increased multimedia capabilities and the broad acceptance of social networks, have allowed an always-on lifestyle for the modern connected individual. The adaption of enriched multimedia content to multi-screen environments, which enables the automatic porting to different target platforms, such as regular web pages, mobile pages and mobile apps, as well as TV applications, is a significant asset for the global market of enriched multimedia content.

Based on the above observations, we introduce the Multimedia Content Annotations for Rapid Exploitation in Multi-Screen Environments (MECANEX) toolkit that consists of the following key components/tools: a) the automatic annotation and editorial tool, b) the multi-screen tool and c) the social and personalization tool. The first tool aims at annotating the multimedia content via adapting appropriate automatic annotation methodologies, algorithms and techniques, where the annotations consist of generic objects, faces and landmarks. Furthermore, via the annotation and editorial tool the MECANEX toolkit offers the ability to semi-automatically enrich video collections for multi-screen applications. The enriched and annotated multimedia content, if stored and organized within archiving systems, may be utilized in order to create and synthesize new multimedia content and advertising campaigns by using already available building blocks. Moreover, the MECANEX toolkit will allow to rapidly prototype multi-screen concepts and test related interaction models, allowing the development of solid business models, marketing and advertising campaigns.

On top of the first two tools, the social and personalization tool will build customers' profiles via collecting information about them implicitly (based on their actions) or explicitly (based on their explicit feedback). Users' profiles allow the improvement of users' Quality of Experience (QoE) by delivering to them more related content to their needs and preferences and will provide feedback to the process of synthesizing new multimedia content and advertising campaigns based on the specific features of the target audience. It should be noted that in this article, though the whole MECANEX toolkit's architecture is presented, special emphasis is given on the social and personalization tool.

Such tools can be potentially adopted by production and post-production companies, as well as by content providers or relevant service oriented companies and will be adapted into the creation process. In particular, the automatic annotation support will lower annotation costs for broadcasters and production houses, will lead to advanced search and retrieval technologies that will reduce the time/cost for finding relevant pieces of content and will improve the quality level of content for re-use. This content although re-used will be adapted to the needs/preferences of each specific audience via the social and personalization tool. Advertising agencies will benefit from t towards designing sophisticated user-engaging content and context-aware personalized advertising services that maximize users' experience and increase brand awareness and conversion rates.

The rest of the paper is structured as follows. Section II presents the overall MECANEX architecture and describes each constituent tool. In the sequel, Section III studies a use case scenario of the MECANEX toolkit for advertising companies, and finally, Section IV concludes the paper.

## II. MECANEX ARCHITECTURE

Fig. 1 illustrates the overall MECANEX architecture where we can distinguish between the backend workflow layer and the separate MECANEX tools communicating through well-defined APIs. The multi-screen tool that lies on top of the basic workflow, allows developers to create multi-screen, multi-user and multi-device type applications. MECANEX follows the lightweight coupling paradigm, thus all separate MECANEX tools are standalone services, language/platform independent, communicating through MECANEX interfaces (REST based) to allow integration over internet based protocols. Two distinct modes of operation are considered, one for the business to customer (B2C) scenarios, where the end-users are simple viewers and one for the business to business (B2B) scenarios, where the end-users are professionals. In the following subsections, the separate tools synthesizing the MECANEX toolkit are presented in more detail.

### A. Automatic Annotation & Editorial Tool

This tool includes the multimedia content analysis, automatic annotation, search and retrieval and editorial operations. One of the main goals of MECANEX is to provide tools and functionalities that enrich multimedia content with useful annotations and metadata. The automatic annotation

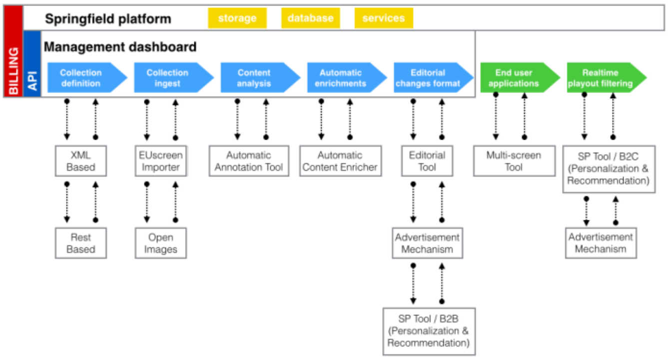


Fig. 1. The MECANEX toolkit's overall architecture

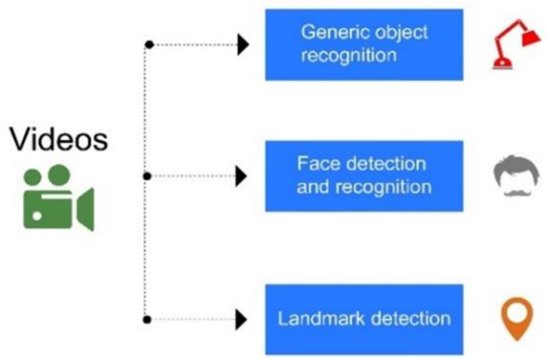


Fig. 2. Automatic Annotation tool

tool provides such functionality without user intervention, by automatically generating annotations on a per frame basis for each video.

- **Multimedia content analysis**

During this phase, which is a prerequisite for the automatic annotation operation, all videos contained in the MECANEX data repository are analyzed for their further processing and creation of annotations. This analysis targets at the extraction of content-related information from multimedia data. The multimodal analysis and description of multimedia content is an active research topic. It is widely accepted that multimodal analysis is a key to achieve a significant increase of robustness and capabilities in multimedia content description [1]. Successful examples are the detection of events in sport videos [2], sport video browsing [3], or the detection of repeating events in multimedia streams [4].

The methodology followed treats video frames extracted at specific time intervals as individual images, and processes them with image analysis algorithms. More specifically, three different image analysis tasks are performed, namely a) face detection and recognition, b) generic object detection and c) landmark detection/recognition. The metadata features extracted are used as supplementary time based metadata in the production process, such as the name, the time stamp, the source of an image or video, as well as the duration and the location of shots. Thus, they can provide extra information to text and visual features, functionality of great importance given the emerging huge collections of digital video (i.e. Digital Video Libraries - DVL).

- **Automatic annotation technology**

The automatic annotation exploits technologies, related to the extracted information regarding the presence of objects, landmarks or persons in images and videos, and supplements them. Video files serve as input and are processed at frame level, where in cases of low image quality individual frames are visually enhanced by appropriate image processing techniques for noise reduction and contrast enhancement. As shown in Fig. 2, the automatic annotation tool consists of three independent modules, working in a parallel fashion and creating a single output file for each analyzed video, containing all the extracted data along with time (timestamps) and place (coordinates) information.

The exploitation of automatically generated metadata for videos, as well as per frame annotations, gives rise to a better user experience by improving the search functionality, as well as providing information in order to produce different kinds of enrichments for the broadcasted content.

- **Search and retrieval**

These functionalities will provide the flexibility to reuse already analyzed, enriched and annotated multimedia content, stored and organized within archiving systems in order to create and synthesize new multimedia content and advertising campaigns via already available building blocks. Semantic search methodologies are supported, i.e. search based on high-level semantic metadata and annotations, giving professional users a high degree of freedom.

- **Editorial tool**

The editorial tool consists of the content enrichment component and the editorial dashboard, and its operation is the automatic inclusion of more semantic to the content extracted in the automatic annotation tool while importing the data to the editorial tool. It also allows the professional users to add or correct enrichments manually through the user interface (UI) of the editorial tool. The content enrichment component uses semantic web technologies to make powerful enrichments and to easily connect to other semantic ontologies. The annotations detected in the automatic annotation procedure serve as input where semantic enrichments, e.g. text, links, video, image, etc., and additional information is added. The editorial tool is based on the non-linear-video (NLV) concept, which provides functions to create, distribute and play interactive video content across platforms and devices by making objects in the video clickable for the viewer.

### *B. Multi-screen Tool*

After the multimedia content analysis, the automatic annotation and the editorial procedures have been realized, the enriched video is stored in the system. However, there still exists the problem of how, when and where to present the enriched content to the end user(s) / consumer(s). The content now has many extra informational hooks that make it easy to show, for example, relevant ads or related content but that all such information is statically presented. At this point, the multi-screen tool provides a decision engine that can take the enriching metadata, advertisements, live detected social data and information of the current context to create a scenario and display it on the available unified screens. Based on the capabilities of the screen the information displayed is adjusted. The multi-screen tool aspires to create a workable framework that can be used to create enriched multi-screen applications in real world scenarios, which implies adaptability of the system due to absence of knowledge of what, when and how users consume. The ultimate goal of the multi-screen tool is to enable easy implementation of prototypes and applications that want to use the MECANEX platform in a multi-device and multi-screen setting. This is achieved initially through the creation and use of easy to use APIs into our SaaS services that interact with the services being able to create new interactive frontends. Furthermore, a frontend framework is provided with easily adaptable pre-built building blocks. These building blocks will use the APIs in the platform and turn them into ready to use features. Examples of these building blocks could be login methods and user profile parts, book-marking, commenting, voting and sharing elements and other parts needed in multi-screen applications. The multi-screen tool is illustrated in Fig.3.

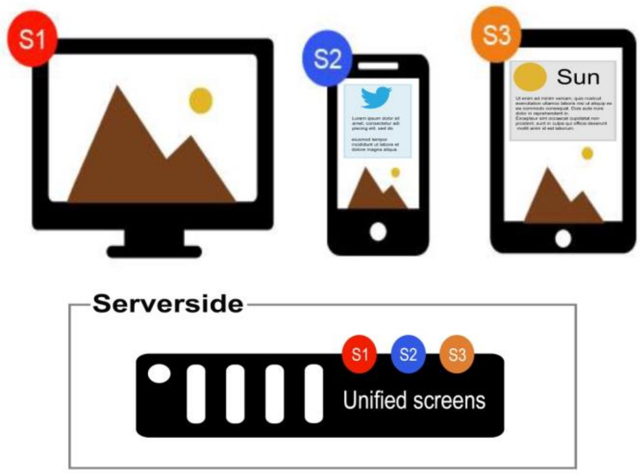


Fig. 3. Overview of the Multi-screen tool

### *C. Social & Personalization Tool*

The social and personalization tool (SP Tool) consists of three main interacting functionalities/mechanisms, namely personalization, relevance feedback and social recommendation as it is illustrated in Fig. 4 [5], [6]. More specifically, each user has the option to provide personal information related either to demographic data and/or personal interests on specified thematic units, while additional information can be extracted by the public profiles in social networks if the latter are provided by the user. This information initializes the personalization mechanism, i.e. the

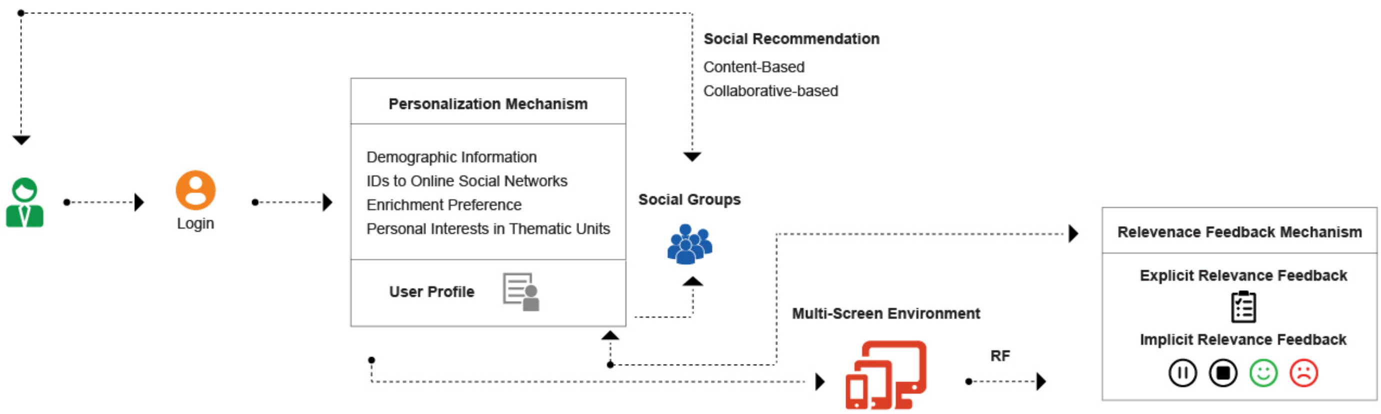


Fig. 4. Illustration of the Social & Personalization tool's operation

user's profile which is then updated via relevance feedback. To achieve this, the relevance feedback mechanism gathers all the explicitly or implicitly provided user feedback via his/her interactions with the multimedia content, thus acting as input to the personalization mechanism. The third mechanism, i.e. the social recommendation one, takes as input the user profiles for providing each user with multimedia content recommendations. These three main mechanisms are efficiently combined towards delivering personalized information to the end-user close to his/her interests, improving end users' Quality of Experience (QoE) and making the overall MECANEX toolkit's usage more appealing.

With respect to the B2C scenarios, SP tool targets at providing viewers with the benefits of personalized experience, i.e. the SP tool recommends content to the end-users based on their preferences. Recommendation is not limited to the video content itself, but expands to the enrichments and advertisements that accompany the video. The SP tool delivers a ranked list of the enriched multimedia content (videos accompanied with enrichments and/or advertisements) to the end-users. With respect to the B2B scenarios, the SP tool creates clusters [7] of the MECANEX users on demand of the professional MECANEX users and also based on the features that the professional users provide as input for the clustering (i.e. demographic users' features). Clustering is performed based on state-of-the art algorithms, e.g. k-means or spectral clustering [7]. Then the SP tool delivers to the professional users each cluster's characteristics along with a list of recommended multimedia content (videos, enrichments, and advertisements) for the corresponding cluster. The importance of the SP tool in the B2B scenario lies in assisting/advising professional users to create/deliver multimedia content targeted to the specific characteristics of their audience. Both scenarios are based on the creation of users' profiles [8], i.e. the personalization mechanism, their update via the relevance feedback mechanism and then their usage from the recommendation mechanism to identify related multimedia content for each user (viewer or professional user) (Fig. 4). Thus, the SP tool mainly creates and updates user profiles capturing user preferences, needs and interests which are used for the recommendation process in a quite different way between the B2B and B2C scenarios as explained above.

Users' profiles are developed based on the Dynamic Planetic User Model [9] that consists of planets, i.e. nodes, and interconnections among planets, i.e. edges. The number of planets for all users' profiles is fixed and matches to the specific set of annotation terms and features whereby the multimedia content has been indexed [10]. An appropriate weight is associated with each planet, as well as with each interconnection among different planets, and higher weight values indicate high importance of single (planets) or pairs (edges) of features for the user. User profile update consists of weights' updates and is performed via relevance feedback i.e. based on user interactions with the multimedia content. Relevance feedback is divided in implicit and explicit one, where implicit relevance feedback [11] consists of observations of user behavior during her/his interaction with the multimedia content (i.e. shares in social media, clicks on enrichment and advertisements, playtime of the video, etc.), and explicit relevance feedback corresponds to users' clear statement of opinion about the multimedia content (Fig. 4). The final user profile used in the recommendation process is a weighted vector derived by the multiplication of the planets' weight vector and their interconnections' weight matrix. The MECANEX platform captures dynamically such user activity, i.e. interaction with the multimedia content, and communicates it to the SP tool where it is translated as feedback to update user profiles.

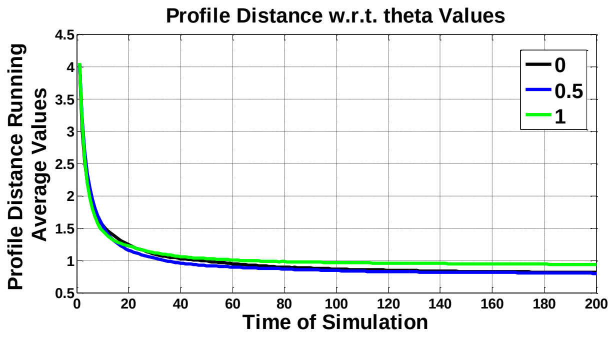


Fig. 5. Indicative evaluation of the SP tool via off-line experimentation

Following, a content-based recommendation mechanism [12] is proposed based on inner product or other similarity measures [13] between user's profile and multimedia content characteristics (where both are described via a specific set of feature terms coinciding with the planets of users' profiles). The main drawback of the content-based recommendation is the overspecialization because by nature it tends to recommend the same types of content. Thus, to overcome this problem a collaborative-based recommendation [14] algorithm is introduced in convex combination with the content-based one. The collaborative-based recommendation is a group's intelligence method, which uses the group's intelligence to solve the problem of single user's personalized information need. The final recommendation algorithm is of hybrid form [15] and is appropriately tuned to each scenario, i.e. B2B, B2C, where for the former clustering methods are also applied as explained above. While the SP tool has been integrated within the MECANEX platform, it can be easily integrated with other social media platforms. SP tool exposes a REST API consumed by the corresponding media platform which then can provide our tool with all the necessary data, i.e. the user personal information, the video ids that the user has selected as well any other user actions perceived as relevance feedback.

In Fig. 5 we present indicative evaluation results of the Social and Personalization tool, performed via off-line simulations. Off-line experimentation aims at evaluating the performance of the SP tool based on simulated user behavior/actions under reasonable assumptions through the definition of user behavioral classes. In these experiments we have *5* user behavioral classes, and in each recommendation round (horizontal axis) we recommend 6 videos to each user from a total video collection consisting of *20* videos classified in *5* thematic classes. Each video has *8* enrichments and *8* advertisements and the user interacts with them and the video itself. Each experiment lasts for *200* recommendation rounds (horizontal axis) during which the user interacts with all the recommended videos and feedback is collected to update his/her profile. The relevance feedback types considered are the percentage of video watched, the number of clicks on enrichments/advertisements, the number of shares of enrichments on social media and the explicit relevance feedback given by the user. Specifically, we examine the euclidean distance between the users' real profiles and the corresponding profiles inferred by the SP tool. "theta" values (mentioned in figure's title and given in the legend) determine the amount of contribution of content and collaborative recommendation in the hybrid scheme where the zero value corresponds to only collaborative filtering and the unity value to only content filtering. Given that the number of features (or planets of the users' profiles) is equal to a hundred, the above defined profile distance attains low values for all values of theta (*0*, *0.5*, *1*), indicating the efficiency of the SP tool. In addition, the participation of collaborative filtering in the hybrid scheme further decreases the distance values.

## III. USE CASE

In order to better demonstrate the effectiveness of the proposed multimedia framework the following use case is presented. The goal is to maximize users' experience by providing enhanced second screen interaction based advertising services. Users, through the MECANEX toolkit, will be the recipients of sophisticated user-engaging content- and context-aware personalized advertisements that increase brand awareness and conversion rate.

Suppose that users are relaxed on their couch watching a documentary featuring the Seven Wonders of the Ancient World. The streaming network offers a second screen application enabling viewers all around the globe to i) browse additional information about the currently displayed content and ii) share their thoughts and opinions with their social networks. Towards offering user-engaging personalized advertising services, viewers receive on their second-screen notifications and motivational messages prompting to accomplish various challenges, solve quizzes and collaborate towards collecting points, participating in leader board games, earning loyalty coupons or even receive bundle offers.

The MECANEX toolkit enables marketing and advertising agencies to achieve campaign optimization and increase conversion rate by exploiting the annotated multimedia content that users are currently watching, not just in a per hour basis (i.e. Seven Wonders of the Ancient World) but on a per second basis (i.e. now watching the Colossus of Rhodes) in order to identify and deliver to user the most relevant advertisement content.

Moreover, via the MECANEX multi-screen API, new applications can be developed dynamically exploiting the annotated multimedia information on the fly, e.g. producing a generic MECANEX-enabled Quiz application and thus maximize user experience (Fig. 6). The users will have the chance to earn points or loyalty coupons by playing educational and entertaining games and participating in challenges e.g. which is the only one of the 7 ancient wonders still standing?

Finally, personalized content delivery is achieved, since behavioral and demographic information, as exposed via the social presence of the viewer as an individual (e.g. posts or comments on the content) or as a group member (e.g. interests of friends, discussions), allows the delivery of customizable personalized promotions, bundle offers, advertisements etc. optimizing the marketing campaign (Fig. 7). The relevance feedback, personalization and social recommendation MECANEX API connects users' responses and comments with movie streams and annotation data, allows access to user preferences and feedback, thus providing clustering information of users and trends within the content lifecycle.

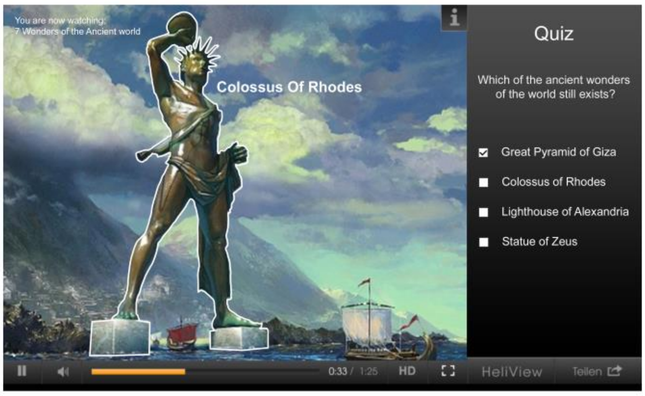


Fig. 6. Example of Quiz application

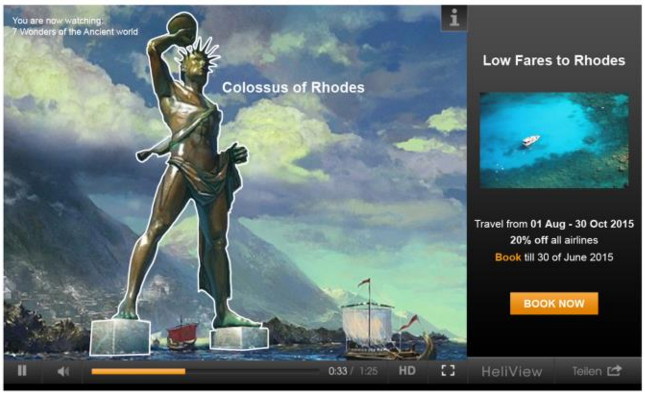


Fig. 7. Example of relevant advertisement content, i.e. Low Fares to Rhodes

## IV. CONCLUSIONS

In this article the MECANEX toolkit was introduced and described for the provisioning of innovative tools for annotation and editorial support of multimedia content, extraction of personalized information and adaption of enriched multimedia content in multi-screen environments. Personalization and recommendation schemes are particularly leveraged towards improving end-users' experience via creating and delivering to them enriched multimedia content aligned to their preferences and needs. The developed toolkit is envisioned to be adopted by production and post-production companies, as well as by content providers or relevant service oriented companies, where its use will improve the chance for SMEs to develop more solid business models, marketing and advertising campaigns. Future work within the MECANEX project includes the integration of the separate tools within existing production infrastructures, and the assessment of the integrated toolkit in terms of system-level effectiveness, performance and fulfillment of user and investor requirements and needs.


## ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 Framework Programme for research and innovation under grant agreement no 645206.


## REFERENCES


[1] P.A. Maragos, A. Potamianos, P. Gros, “Multimodal Processing and Interaction: Audio, Video, Text”, Springer, 2008.

[2] S.-C. Chen, M. Chen, C. Zhang, M.-L. Shyu; “Exciting Event Detection Using Multi-level Multimodal Descriptors and Data Classification”, Proc. of the 8th Int’l Symposium on Multimedia, pp. 193-200, 2006.

[3] H. Y. Liu, H. Zhang, “A Sports Video Browsing and Retrieval System based on Multimodal Analysis: SportsBR”, Proc. of the Int’l Conf. on Machine Learning and Cybernetics, pp. 5077-5081, 2005.

[4] C. Herley, “ARGOS: Automatically Extracting Repeating Objects from Multimedia Streams”, IEEE Trans. on Multimedia, vol. 8, no. 1, pp. 115-129, 2006.

[5] J. Bobadilla, F. Ortega, A. Hernando, A. Gutierrez, “Recommender Systems Survey”, Knowledge-Based Systems, Elsevier, Vol. 46, pp. 109-132, April 2013.

[6] C. Buckley, G. Salton, J. Allan, “The Effect of Adding Relevance Information in a Relevance Feedback Environment”, Proc. of the 17th Annual Int’l ACM Conference on Research and Development in Information Retrieval (SIGIR '94), Springer-Verlag New York, pp. 292-300, 1994.

[7] D. Yan, L. Huang and M. I. Jordan, “Fast Approximate Spectral Clustering”, Proc. Of the 15th ACM Conference on Knowledge Discovery and Data Mining (SIGKDD), pp. 907-916, 2009.

[8] S. Gauch, M. Speretta, A. Chandramouli, A. Micarelli, “User Profiles for Personalized Information Access”, pp. 54-89, Book: The Adaptive Web, Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2007.

[9] G. Gentili, A. Micarelli , F. Sciarrone, “Infoweb: An Adaptive Information Filtering System for the Cultural Heritage Domain”, Applied Artificial Intelligence, Vol. 17, No. 8-9, pp. 715-744, 2003.

[10] H.C. Wu, R.W.P. Luk, K.F. Wong, K.L. Kwok, “Interpreting TF-IDF Term Weights as Making Relevance Decisions”, ACM Transactions on Information Systems, Vol. 26, No. 3, 2008.

[11] F. Hopfgartner, J. Jose, “Evaluating the Implicit Feedback Models for Adaptive Video Retrieval”, Proc. of the Int’l Workshop on Multimedia Information Retrieval, pp. 323-331, September 2007.

[12] R. Meteren, M. Someren, “Using Content-Based Filtering for Recommendation”, in Proc. of ECML Workshop: Maching Learning in Information Age, pp. 47-56, 2000.

[13] S.-H. Cha, “Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions", Int’l Journal of Mathematical Models and Methods in Applied Sciences, Vol. 1, No. 4, pp. 300-307, 2007.

[14] B. Schafer, D. Frankowski, J. Herlocker, S. Sen, “Collaborative Filtering Recommender Systems”, The Adaptive Web Chapter 9, Springer, pp. 291-324, 2007.

[15] C. Porcel, A. Tejeda-Lorente, M. A. Martinez, E. Herrera-Viedma, “A hybrid Recommender System for the Selective Dissemination of Research Resources in a Technology Transfer Office”, Information Systems, Vol. 184, No. 1, pp. 13-32, 2012.