

euTV: a System for Media Monitoring and Publishing

Marco Bertini, Alberto Del Bimbo
Università degli Studi di Firenze - MICC
Firenze, Italy

Emile Bijk
Hogeschool voor de Kunsten Utrecht
Hilversum, The Netherlands

George Ioannidis
IN2 search interfaces development
Edinburgh, United Kingdom

Isabel Trancoso, Hugo Meinedo
Inesc-ID
Lisbon, Portugal

ABSTRACT

In this paper, we describe the euTV system, which provides a flexible approach to collect, manage, annotate and publish collections of images, videos and textual documents. The system is based on a Service Oriented Architecture that allows to combine and orchestrate a large set of web services for automatic and manual annotation, retrieval, browsing, ingestion and authoring of multimedia sources. euTV tools have been used to create several publicly available vertical applications, addressing different use cases. Positive results of user evaluations have shown that the system can be effectively used to create different types of applications.

Categories and Subject Descriptors

H.3.5 [Information Storage and Retrieval]: Online Information Services; H.4 [Information Systems Applications]: Miscellaneous

General Terms

Algorithms, Design, Experimentation

Keywords

Semantic multimedia annotation; SOA; content-based multimedia retrieval

1. INTRODUCTION

The explosion of digital data in recent times, in its varied forms and formats (H.264 and Flash videos, MP3 and Wav audio, HTML and PDF documents, images), requires the creation of effective tools to organize, manage and link digital resources, in order to maximize accessibility and reduce cost issues for everyone concerned, from content managers to online content consumers. On a larger scale, isolated multimedia repositories developed by content owners and technology providers can be connected, unleashing opportunities for innovative user services and creating new business models, in the vein of on-demand, online or mobile TV ventures.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

MM'13, October 21–25, 2013, Barcelona, Spain.

ACM 978-1-4503-2404-5/13/10.

<http://dx.doi.org/10.1145/2502081.2502265>.

The euTV¹ project stems from above conditions and potentialities to connect publicly available multimedia information streams under a unifying framework, which additionally allows publishers of audio-visual content to monetize their products and services. The backbone of euTV is a scalable audio-visual analysis and indexing system that allows detection and tracking of vast amounts of multimedia content based on Topics of Interest (TOI), that correspond to a user's profile and set of employed search terms. The front-end is a portal that can be customized to be suitable to different scenarios and scopes, displaying syndicated content, allowing users to perform searches, refine queries, and produce faceted presentation of results.

2. DEMO

The demo shows three different applications created using the euTV media monitoring and publishing platform:

- FilePanel – files your way²: a web-based media management system for individuals and groups (private and institutional).
- NewsBoard – news at your fingertips³: A board to collect news feeds, process them, and filter and search what is interesting to the end-user. This application targets mainstream media clipping and monitoring markets.
- TravelPulse – photos in a pulse⁴: The application targets professional content providers, notably touristic destinations managers supporting them in promoting their cities/regions. It has been conceived as a showcase of the euTV framework for tourism applications.

3. THE SYSTEM

The euTV platform has a powerful infrastructure, that embraces different programming and scripting languages (C++, Java, PHP, Javascript, LUA), required for the many euTV tools and services, and libraries and open source software products (FFmpeg, Red5, Mule etc.) to process audio, visual and textual multimedia documents. The projects scale has required a pragmatic and robust approach to integrate all components into a scalable platform. This goal has been reached by extensively using a services oriented approach for system integration, with an agnostic approach to languages

¹<http://www.eutvweb.eu> - The research leading to these results has received funding from the EU 7th FP managed by REA (FP7/2007-2013) under grant agreement no. 262428.

²<http://eutvweb.eu/filepanel/>

³<http://eutvweb.eu/newsboard/>

⁴<http://eutvweb.eu/travelpulse/>

and protocols. The platform has a central services-based architecture layer consisting of a file-store, a repository and a web services hub (Apache Mule). This layer acts as the central hub to which the media processor framework and the end-user interfaces could rely on for storing and retrieving data in a flexible manner (Fig. ??).

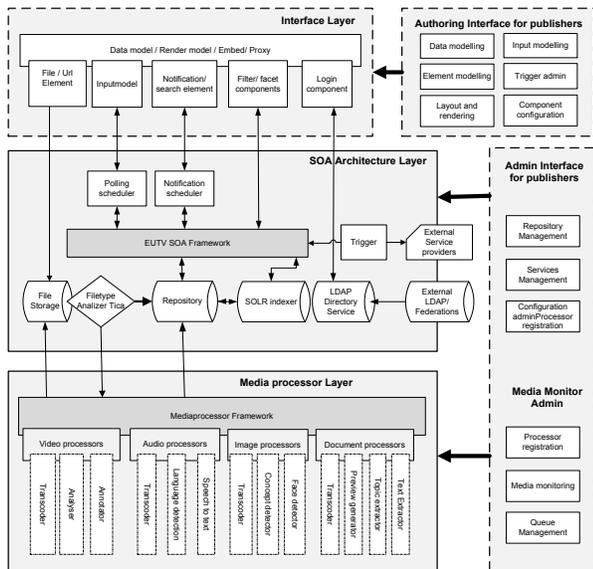


Figure 1: euTV system architecture.

Text indexing and annotation is performed by services that provide: *i*) language classification, based on n-grams and Naïve Bayes classifiers that despite the simplicity have shown to work effectively also on short fragments [?]; *ii*) topic detection based on LDA; *iii*) named entity extraction based on gazetteers and a rule-based system, to handle entities that have not been added yet to lists [?]. Topic detection and named entity identification can be used also with the outcomes of speech transcription services.

Speech and audio indexing and analysis tools provide services for: *i*) audio segmentation [?] that separates audio streams into six separate components: four for classification (speech / non-speech, gender, background and speaker identification), one for speaker clustering and one for acoustic change detection. These components are mostly model-based, making extensive use of feed-forward fully connected Multi-Layer Perceptrons trained with back-propagation; *ii*) audio language identification, based on [?], that identifies the 12 most spoken languages across the European Union, using SVMs for the phonotactic system and an I-vector based acoustic sub-system; *iii*) an audio event module that recognizes 54 sound concepts, using a combination of MFCC, ZCR and MPEG features to feed SVM classifiers; *iv*) audio transcriptions, an engine that uses a hybrid approach combining the temporal modelling capabilities of Hidden Markov Models with the pattern discriminative classification capabilities of Multi-Layer Perceptrons [?], and works with English, Spanish, Portuguese and German.

Visual annotation and indexing deal with images and videos at syntactic and semantic levels. Similarity-based retrieval deals with images and video keyframes, using a combination of MPEG global features and SIFT descriptors, indexed using approximate similarity searching based on inverted files [?] for scalability. Semantic annotation is obtained using a BoW-based approach, with a model selection step to select

the best combination of interest point detectors/descriptors for each concept classifier, and using the PMK kernel [?] for efficient computation. Classifiers can be trained with a specific service, that exploits social media as training source [?].

Each type of media has a specific component that can handle visualization, media search and manual annotation (Fig. ??). A web-based authoring system allows to combine all the services to design specific applications for each use scenario, from the automatic ingestion of media to their processing, search and presentation (Fig. ??).

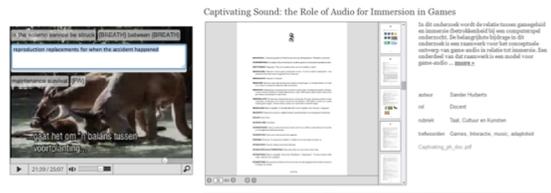


Figure 2: left) video play/annotations, showing automatic transcriptions; right) document viewer/annotator, showing concepts and entities on the right.

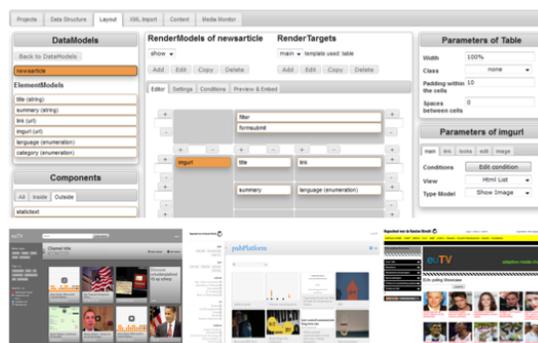


Figure 3: top) euTV system to design application interfaces; bottom) three examples of euTV applications.

4. REFERENCES

- [1] H. Zhang. The optimality of Naïve Bayes. In *Proc. of FLAIRS2004*, 2004.
- [2] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proc. of ACL*, 2002.
- [3] H. Meinedo. *Audio pre-processing and speech recognition for broadcast news*. PhD thesis, IST, Lisbon, 2008.
- [4] A. Abad. The l2f language recognition system for albayzin 2012 evaluation. In *Proc. of Iberspeech*, 2012.
- [5] J. Neto, H. Meinedo, M. Viveiros, R. Cassaca, C. Martins, and D. Caseiro. Broadcast news subtitling system in portuguese. In *Proc. ICASSP*, 2008.
- [6] G. Amato and P. Savino. Approximate similarity search in metric spaces using inverted files. In *Proc. of InfoScale*, 2008.
- [7] K. Grauman and T. Darrell. The pyramid match kernel: Efficient learning with sets of features. *Journal of Machine Learning Research*, 2007.
- [8] A. Ulges, C. Schulze, M. Koch, and T. Breuel. The challenge of tagging online video. *Computer Vision and Image Understanding*, 2009.