

# MEDIACAMPAIGN – A MULTIMODAL SEMANTIC ANALYSIS SYSTEM FOR ADVERTISEMENT CAMPAIGN DETECTION

*Herwig Rehatschek<sup>1</sup>, Robert Sorschag<sup>1</sup>, Bernhard Rettenbacher<sup>1</sup>, Herwig Zeiner<sup>1</sup>, Julien Nioche<sup>2</sup>, Franciska DeJong<sup>3</sup>, Roeland Ordelman<sup>3</sup> and David van Leeuwen<sup>4</sup>*

<sup>1</sup>JOANNEUM RESEARCH Forschungsgesellschaft mbH, Steyrergasse 17,  
A-8010 Graz, Austria

{Herwig.Rehatschek, Robert.Sorschag, Bernhard.Rettenbacher, Herwig.Zeiner}@joanneum.at

<sup>2</sup>University of Sheffield, West Bank,  
Sheffield, United Kingdom

J.Nioche@Sheffield.ac.uk

<sup>3</sup>University of Twente, Drienerlolaan 5,  
7522 NB Enschede, The Netherlands

{fdejong, ordelman}@ewi.utwente.nl

<sup>4</sup>TNO Human Factors, Postbus 23,  
3769 ZG Soesterberg, The Netherlands  
david.vanleeuwen@tno.nl

## ABSTRACT

MediaCampaign's scope is on discovering and inter-relating advertisements and campaigns, i.e. to relate advertisements semantically belonging together, across different countries and different media. The project's main goal is to automate to a large degree the detection and tracking of advertisement campaigns on television, Internet and in the press. For this purpose we introduce a first prototype of a fully integrated semantic analysis system based on an ontology which automatically detects new creatives and campaigns by utilizing a multimodal analysis system and a framework for the resolution of semantic identity.

## 1 INTRODUCTION

More than 70 billion Euros are spent each year in the EU advertising sector which makes it a huge market segment. The independent measurement of advertisement expenditure is performed by media monitoring companies who investigate advertisements on a daily basis in different countries and media. This information is usually collected manually in the media, Press, TV, Radio, Internet, Cinema, Outdoor, and others, down to the product level. This means that for each day it is known how much money a company spent on one specific product in one media, such as Press, in a country. Obviously this information is very important for executives and decision makers being used as a basis for marketing decisions. As an example, imagine for the IT

sector the introduction of a new laptop. The marketing manager of the company needs to know how much the competitors have invested in advertising in his segment, in order to bring a similar product successfully into the market. Another important example would be sponsorship tracking. This is where companies want to measure the effectiveness of their advertisements on hoardings, players, cars, etc., in comparison to having invested their money directly into advertisements in Press, TV, Radio and Internet.

The collection process is very complicated and mostly performed manually. This is especially true for the media Press. The media Radio and TV are at some stages already supported by technology. The complexity is driven by the enormous diversity of advertisements, the range of media (Press, TV, Radio, Internet, Cinema, outdoor), and the different languages involved.

Within the project MediaCampaign we cope with this complexity. However, rather than having the quite unrealistic goal of automating the entire media monitoring sector, taking into account all the various existing and future business cases, we concentrate on one specific area within the domain. Within MediaCampaign we want to discover advertisement campaigns. An advertising campaign is defined as a collection of advertisements semantically belonging together, across different countries (NL, D and UK are covered in the project – hence the system supports textual and speech analysis in the languages Dutch, German and English) and different media. The project's main goal is

to automate to a large degree the detection and tracking of media campaigns on Television, Internet and in the Press.

The main scientific challenges in this process involve the detection of new advertisements (called 'creatives' in business jargon) and the interrelation of advertisements belonging semantically together to campaigns. The new creative detection is based on the fusion of a multi-modal semantic analysis for the modalities image, video, audio and text. The knowledge fusion and campaign detection is based on relating metadata, initially acquired during the processing of spots (incoming new advertisements not classified yet), and the formal model. This represents both campaigns and creatives metadata on their properties and relations as they are defined in Media Presence and Campaign Ontology (MEPCO) [1]. Hence the main workflow steps involved in the MediaCampaign project are data acquisition, multi-modal analysis, creative detection, knowledge fusion and campaign detection, and finally the delivery system for querying and displaying results.

This paper describes the system architecture and the main technical workflow of the MediaCampaign prototype system (manifested in chapter 3) together with some of the main scientific innovations of the MediaCampaign project achieved so far. In detail the most current advances in the multi-modal semantic analysis is given in Chapter 4. Chapter 5 is dedicated to the knowledge fusion, creative and campaign detection. The last chapters in this paper deal with conclusions and future work to be performed.

## 2 RELATED WORK

### 2.1 Visual analysis

There are semi-automatic commercial solutions and ongoing research projects in existence, only for dedicated business cases in the advertising analysis and media monitoring domain. For the very specific (unimodal) task of brand detection the R&D project DETECT and commercial products from Sportsi, Spikenet Technology, and Omniperception can be identified. Furthermore it has to be mentioned that there are many service providers in the market, however, very few technological solution providers. Especially in the commercial area we are not aware of any system which uses a multimodal approach in order to extract fused semantic information.

DIRECT-INFO [2], [3] was finalized in March 2006 and was specifically targeted on sponsorship tracking within TV broadcasts by utilizing a distributed multimodal analysis system. MediaCampaign will be partially based on DIRECT-INFO results.

DETECT is a completed EC project which aimed to automatically extract features from digital video streams. Based on the extracted features, DETECT automatically generated statistics. In particular logos were extracted and frequency statistics were generated [4].

In the area of TV fingerprinting and commercial block detection, research results exist from [5], [6], and [7]. [5] describes an algorithm for redetection of advertisements within a TV stream with a precision of 99% and a recall rate of 95%. This approach is mainly based on an acoustic matching method while visual descriptors are used only for verification. In this work redetection of advertisements (ads) is performed to replace one ad with another ad for personalisation purposes. In contrast, in MediaCampaign detection/clipping of ads is used to find out if they are new or not, and in order to calculate different statistics and key data. [6] describes a classification approach based on SVM in order to find commercial blocks within video streams. In MediaCampaign, we use a similar classification method although it is used for another task. [7] was a diploma thesis performed within DIRECT-INFO project. It proposes a method for semantic classification of video that extracts a set of basic features such as logo present or not, split screen, dominant color, etc. from the signal. These features are then used to classify the content into some predefined genres (commercial, comic, news, sports, etc.). The concept of genres has been kept generic. As MediaCampaign is aimed on the detection of new ads (or media campaigns), segmenting the input into single advertisements, is not a topic in MediaCampaign.

### 2.2 Textual analysis

Metadata obtained automatically from textual analysis of multimedia content have been used in a number of projects, in particular for indexing and retrieving purposes.

The MUMIS project [8] provides technology for indexing and retrieval of Multimedia Programs in the domain of soccer in multiple languages. Information Extraction components for each language obtain key entities and events from different sources: structured documents, news, spoken and transcriptions. The partial information extracted from each individual source is then combined, which involves the disambiguation of the entities found (cross document co-reference) and the temporal alignment of the scenes among the different documents.

The PrestoSpace<sup>1</sup> project aims at providing services for the preservation of audiovisual content. In PrestoSpace, automatic transcripts of news broadcasts are obtained with Automatic Speech Recognition (ASR) technology. Subsequently, during textual analysis, these transcripts are annotated with regard to the PROTON<sup>2</sup> ontology. The MEPCO ontology used in MediaCampaign is itself an extension of PROTON. The mechanism for content augmentation in PrestoSpace (which leverages on the KIM<sup>3</sup> platform) from the web [9] improves the performance of the semantic indexing on degraded text, such as ASR

---

<sup>1</sup> <http://www.prestospace.org>

<sup>2</sup> <http://proton.semanticweb.org>

<sup>3</sup> <http://www.ontotext.com/kim/>

transcripts. The semantic annotations of the documents are then indexed and help retrieving news broadcasts. Information extracted from text is not used only for indexing and retrieving multimedia documents but also provides valuable non topical information. This is exemplified in the Direct-Info project, which dealt with the multimodal analysis of audiovisual documents and aimed at identifying sponsors brands in sporting events. The textual analysis in Direct-Info was in charge of determining the sentiment associated to an occurrence of a brand (e.g., positive vs. negative) in OCR transcripts and plain text documents. The detection of the brand itself was done by the analysis of the non-textual, visual content of the document.

### 2.3 Audio Analysis

MediaCampaign audio analysis contributes to campaign detection. It does this by automatically generating metadata for Audio Visual (AV) material via the classification of audio at a conceptual level. It segments, labels segments (e.g., speech, music, jingle, laughter, etc.) and yields transcription of speech fragments. It will build on existing insights and technologies gained for a number of subtasks: audio segmentation and classification, ASR, more specifically, Large Vocabulary Continuous Speech Recognition (LVCSR) and word spotting. Research for these topics has been carried for more than decade. The US National Institute of Standards and Technology<sup>4</sup> (NIST) has been supporting several international benchmark activities that have contributed to baseline technologies for segmentation and Spoken Document Retrieval (SDR). This offers a starting point for the domain adaptation and optimization efforts planned for MediaCampaign.

For ASR, recognition modules are typically based on a combination of acoustic models for the recognition of sounds, and language models for the recognition of words in context. Both kinds of models require the availability of large amounts of training data. For acoustic modeling the most successful systems are based on 100 hours or more of annotated speech. Whilst for the language models, the bulk of the training data consists of text corpora of hundreds of millions of words [17].

Performance levels that are state-of-the-art for large vocabulary data, e.g., Broadcast News (BN) speech, are between 10-20% Word Error Rate, depending on the language, type of speech (spontaneous interviews vs. read news transcripts), and the audio quality (noisy telephone speech vs. clean studio recordings). For applications where ASR is not meant to yield near-perfect readable transcriptions, such as audio search, these performance figures have proven to be adequate. SDR has been shown to be robust against Word Error Rates (WERs) of even up to

50% [16]. For other domains than BN, figures below 50% WER are normally harder to obtain. The recognition of speech present in commercials is an outstanding example of a very difficult speech recognition task. Speech in commercials is very atypical and variable. The messages or scripts are usually 'read' by a voice actor (hired because of his/her voice that matches the image of the product) but the speech is different from the type of 'read speech' ASR systems are familiar with. The 'message' can be shouted, sung, whispered, etc. Moreover the presence of commercial-specific vocabulary, such as brand names, may require language model adaptation. Adaptation and tuning is labor-intensive, and relies on the availability of training data and/or available information sources, such as manually generated metadata. Commercially available tool boxes typically come as a black box without options for flexible adaptation. Therefore, we choose to use non-commercial ASR systems in Media Campaign.

Jingle recognition results in another potential cues for detecting campaigns. The jingle recognition task can be seen as an audio identification task. Typical applications for audio identification systems are song identification to find out the performer of the song and audio watermarking for dealing with copyright issues. Research work for audio identification is done mainly in the music information retrieval area. Most music identification systems use content based audio identification. A review can be found in [14]. Besides music identification, musical similarity for finding song structures or detecting cover versions (e.g., [15]) is also relevant. The main difference between music identification and jingle recognition is that music identification is performed on 6-100 seconds of audio data, whereas most jingles are 2-3 seconds of length. For jingle recognition, only a few publications are available. In [13], jingle candidates are found by a classification system and afterwards, an identification system tries to match the candidates by calculating distance measures. Focusing on the media monitoring and broadcast market, audio identification is used for jingle and advertisement recognition as part of complete media watching systems (e.g., Idioma<sup>5</sup>) and broadcast solutions (e.g., MediaFabric<sup>6</sup>).

## 3 MEDIACAMPAIGN SYSTEM ARCHITECTURE & WORKFLOW

The MediaCampaign prototype system architecture is visualized in Fig. 1. and consists of the four main components: 1. media acquisition, 2. multi-modal media analysis, 3. knowledge fusion and 4. campaign discovery; the delivery system and a knowledge store comprised of three components.

---

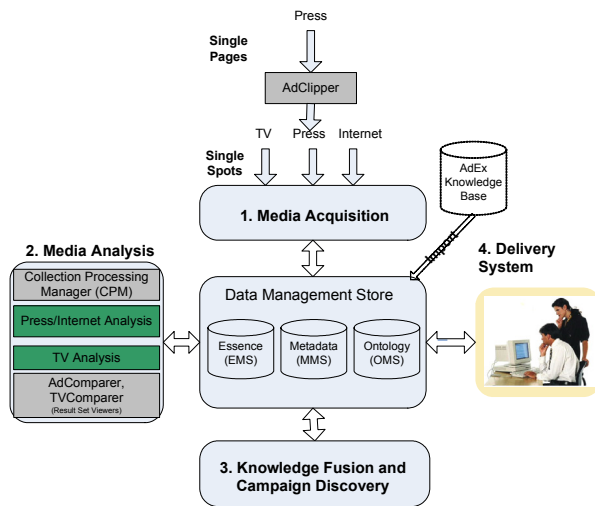
<sup>4</sup> <http://www.nist.gov/>

<sup>5</sup> <http://www.idiomasonline.com>

<sup>6</sup> <http://media.vcs.de>

In the context of the project, the following terms are frequently used: spot, creative and campaign. In order to be clearly understood by the reader we give a short definition of these terms. A creative represents all the different occurrences of a similar spot. However, a creative cannot occur cross media (e.g., TV and press) and cross languages. A spot is a specific, single occurrence of a creative, e.g. a TV spot seen on a given channel at a specific time. The incoming material obtained from the Media Acquisition step is initially available as a spot and later attached to an existing or a new creative (depending whether the spot is already contained in the MediaCampaign database or not). A campaign represents a number of creatives semantically belonging together. A campaign has certain duration and can be cross country and cross media. A campaign may be cross media (e.g., press and Internet) and cross language.

**Media Acquisition:** The acquisition will be carried out on a (potentially) distributed system. The component centralizes the acquisition of all media essences and decouples the system from the different media sources. The acquisition component will accept media from TV and press digital file input (single creatives), for Internet an appropriate grabber will be provided.



**Fig. 1.** Basic system architecture of the MediaCampaign system prototype

**Multi-modal media Analysis:** The goal of the media analysis block is to detect new creatives within a single media and a single country. The analysis subsystem consists of the Collection Processing Manager (CPM), the individual analysis modules, and the result set viewers (AdComparer, TVComparer) for the analysis subsystem. The Collection Processing Manager (CPM) is responsible for managing the entire workflow of the analysis modules covering the media press, Internet and TV, and the modalities audio, video, images and text. Based on the extracted data a fully automatic check is performed whether the creative is new or

existing. The results are presented to an operator, who can manually accept and/ or change the automatic results. After the manual check the creatives are passed to the knowledge fusion and campaign discovery block.

**Knowledge Fusion and Campaign Discovery:** Once all the modalities of an item (a spot) have been analyzed (sound – image – video - text), the creative is compared to existing media campaigns using the information defined in the MEPCO ontology and stored in the Ontology Management Store (OMS). A given creative found in some media, such as press, and published in a given country will possibly be linked to a related creative published in a different media and in a different country.

**Delivery System:** The delivery component presents the overall results of the system. The delivery system focuses on campaigns and provides the results of the cross media interlinking analysis to the operator. The aim of the delivery system is to browse, check, and validate data with respect to existing media campaigns in order to present them to final customers. Customer delivery could be performed by an external system.

**Knowledge store:** The knowledge store consists of the Essence Management Store (EMS), the Metadata Management Store (MMS), and the Ontology Management Store (OMS). The MMS serves as the main exchange point between the analysis systems and to store all low-level analysis data in a static metadata repository. The MMS is used by the creative detector (TV and Press/Internet) in order to decide if a spot is new or existing.

The next chapters give more details on those components of the MediaCampaign system where the most important scientific innovations have been elaborated.

#### 4 SEMANTIC MULTI-MODAL ANALYSIS MODULES

Multimedia content is composed of different modalities, such as video, audio and text. These modalities have to be analyzed in order to extract the most relevant, semantic information for each advertisement. In this context relevant information is produced and company names from the advertisement provide a textual description of the ad. The knowledge of an advertisement is already shown/ printed before it is designated a new advertisement. The types of modalities contained in an advertisement and their relevance depends on the media type (TV, Press or Internet) and the content of the advertisement itself.

The process of analyzing different modalities of an advertisement is not a sequential workflow but a more complex one, because some analysis modules work using the results of other analysis modules. For instance, the text analysis module uses text transcripts given from OCR and ASR analysis modules. The Collection Processing Manager (CPM) is responsible for organizing this workflow appropriately. The most important analysis modules of the

MediaCampaign system are described according to the modality they are analyzing. The modules used to fuse and further process the independent results of these analysis results is described in section 5.

#### 4.1 Visual Analysis Modules

**Image Fingerprinting.** Advertisements in press and internet usually consist of a single image which can include text, such as the company and product brand, a slogan or general product information. In cases where no text exists, the visual similarity is the only modality that can be used to extract information from an investigated ad. Image fingerprinting is done to find existing ads which are identical or similar to the investigated ad. Therefore two steps are performed called similarity matching and exact matching. In the similarity matching step a fast comparison of the investigated ad against all existing ads is performed to generate a list of the most similar ads. The exact matching step is done to find out which of these ads is identical to the investigated ad.

The similarity matching is done with MPEG-7 color descriptors [18] and SIFT-like (Scale Invariant Feature Transform) local image features. First MPEG-7 descriptors are used to select a number of existing ads from the database that roughly consists of the same colors as the investigated ad. Although this method usually returns identical ads, it also returns ads which belong to another company/product and have nothing in common with the investigated ad except the used colors. Therefore the initial similarity matching results are ordered according to textural similarity with the help of SIFT-like features. SIFT is a state of the art object recognition method proposed in [19].

In the exact matching stage each ad found in the similarity matching is individually compared to the investigated creative in descending order. For the individual comparison the two images are first normalized to the same size. Next a local motion registration [[20]] is performed to overcome slight formatting differences of identical ads. At last the two images are matched by a pair wise comparison of 32 squared subregions with the help of gradient histograms with 8 bins.

**Video Fingerprinting.** Video fingerprinting is similar to image fingerprinting for TV ads. It is used to recognize identical and similar ads by a visual comparison of an investigated ad against the existing ones. Video fingerprinting uses a similar matching technique from image fingerprinting on primary extracted keyframes. Each keyframe of an investigated ad is individually matched against all the keyframes of existing ads. A score is derived for the existing ad that belongs to the best matching keyframe. After all keyframes are matched, the existing ads with the highest score are further matched against the input ad according to their shot structure. Therefore the shot length and arrangement is compared individually. Identical

creatives have almost the same shot structure; they can only differ in a few appended shots at the beginning or the end.

**Logo Recognition.** In the TV analysis workflow, logo recognition is done to find out which product or company an investigated ad belongs to. This is especially important for new ads because the logo of a product/ company shown in new ads are usually the same as the logo used in existing ads for the same product/ company. Even though the remaining visual content of the new ad is completely different to the existing one. The logo recognition works with a number of previously learned logos (one or several images per logo) and tries to recognize these logos in a video. It is possible to recognize learned logos at different sizes, rotated, with minor color changes and under different illumination conditions. For this object recognition problem SIFT features [[19]] are used and temporal information of the video is used for verification. As result the size, position and orientation of a recognized logo in the video is generated for each frame of the investigated ad.

#### 4.2 Audio Analysis Modules

**ASR.** ASR is performed with the SHoUT speech decoding system<sup>7</sup>. The speech-to-text module will output phone aligned word strings (first best) and word lattices. The Viterbi search of the decoder has been implemented using the token passing paradigm. Hidden Markov Models (HMM) with three states are used, as well as Gaussian Mixture Models for the probability density functions to calculate acoustical likelihoods of context-dependent phones. Fourgram and trigram backoff language models are used to calculate language probabilities.

The baseline MediaCampaign system will be based on training data for the broadcast news domain. In order to adapt to the MediaCampaign task domain, strategies to adapt to the acoustic diversity in the domain are explored. Optimal domain coverage will require tuning of the recognition vocabulary to list of names of brands and products.

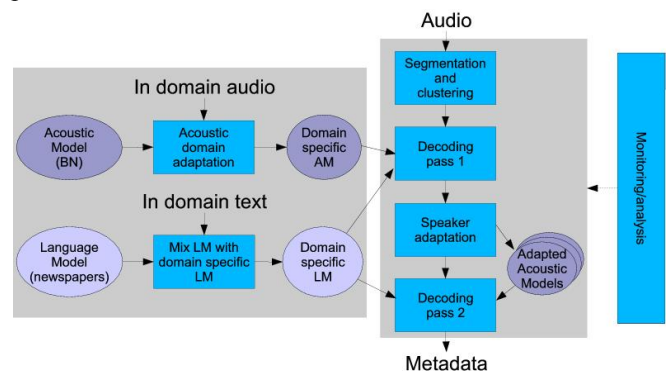


Fig. 2: ASR system set-up

<sup>7</sup> SHoUT is an acronym for Speech recognition University of Twente

Fig. 2 pictures the global ASR system setup. The grey area at the left represents the sub-system responsible for adapting the baseline, broadcast news (BN), acoustic models and language models to the MediaCampaign task domain. The second grey area represents the main audio processing stream. The sub-system consists of four modules. In the first module, the audio is cut up in smaller segments and each segment is assigned to a unique speaker (so-called speaker diarization). After diarization, the second module performs a first speech recognition pass for each speaker using the domain specific models. This recognition result is then used in the third module to create a new acoustic model (AM) for each speaker. These speaker-specific models are eventually used in the fourth module doing the final decoding pass.

**Word spotting.** MediaCampaign also investigates the potential of word spotting approaches. Word spotting may serve as a stand-alone annotation tool, by spotting occurrences of specific names of products and brands in the audio, commercials can be labeled accordingly. Potentially this functionality can help transcription generation via ASR: just like broadcast news topic-specific language models can help to improve ASR performance, brand-specific language models may do so in the MediaCampaign domain. Reversely, word spotting can be deployed *after* transcription generation, to support the spotting/identification of spoken products and brands names in a commercial that have been missed by the ASR, for instance because they are out-of-vocabulary. If a speech transcript is classified as being most probably, about cars, while no car brand name occurs in the transcript, a word spotting step could still try to detect a car brand.

To enable experiments with word spotting functionalities, in MediaCampaign a classical word spotting system is built which is based on a set of acoustic models for the basic phones of a language. A "grammar" is constructed such that at each point in time the recognizer can choose between (i) any of the phone models, which act as fillers, or (ii) a model for the key word. When the path of a keyword is "followed", all the constituent phones in that word are sequentially "matched" with the audio signal. Technically this "choice/ matching" is based on likelihoods where the optimum path is found via the Viterbi algorithm. The engine that is deployed is called SpeechMill and is based upon Sonic technology.<sup>8</sup>

**Jingle Recognition.** Jingles are memorable audio sequences or melodies. Most jingles in TV commercial spots are 2 to 3 seconds in length. In analogy to logo recognition, jingle recognition tries to extract jingles of the audio stream of a TV commercial spot. As jingles are product and company specific, jingle recognition gives additional information. The jingle recognition task can be seen as a fingerprinting

task in the audio-domain, where reference jingles are matched with the spot to be analyzed. As jingle melodies often appear in different pitch, tempo and timbre, the jingle recognition has to be independent to these variations.

The used identification approach for jingle recognition is based on the extraction and matching of a sequence of abstract audio events. We investigate a combination of two different approaches for audio event extraction. First, chroma features, synchronized to the musical beat, provide audio events comparable to musical notes, but are immune to pitching, timbre and tempo variations. Second, more general features like the zero crossing rate, spectral descriptors or Mel Frequency Cepstral Coefficients are extracted from the audio data and the audio events are modeled with Hidden Markov Models (HMM). The audio events are trained unsupervised on short homogeneous segments of a large audio. The audio events extracted from reference jingles are stored in a jingle database and matched with the audio events extracted from the spot to be analyzed.

### 4.3 Text Analysis Module

The Text Analysis module in MediaCampaign analyses text transcripts coming from the OCR and ASR modules and identifies instances of the MEPCO ontology[1], in particular of the classes Advertiser and Product. The Text Analysis is based on GATE<sup>9</sup> [10]. GATE (General Architecture for Text Engineering) is a Java Open Source framework maintained by the University of Sheffield. The system is bundled with components for language analysis, and is in use for Information Extraction (IE), Information Retrieval (IR), Natural Language Generation, summarisation, dialogue, Semantic Web, Knowledge Technologies and Digital Libraries applications. The Text Analysis module utilizes some of the GATE standard resources, such as gazetteers, part of speech tagger and JAPE rules. Additional resources have been developed especially for MediaCampaign, in particular lists of known entities generated from the Adex knowledge base provided by the MediaCampaign industrial partner Nielsen Media Research. This allows the recognition of entities known in the knowledge base but also of new entities, thanks to contextual rules. The Text Analysis module also contains resources for dealing specifically with degraded content, as the incoming documents provided by the OCR and ASR modules. The Text Analysis module combines the information about the frequencies of the entities found in the document and their provenance to rank them at the document level. A future version of this module will use the relations between instances of Advertiser and Product, for instance the knowledge that the product *C3* is related to the company *Citroën*, in order to improve the scoring of the entities. The current version of the Text Analysis module

<sup>8</sup>

[http://slr.colorado.edu/beginweb/speech\\_recognition/sonic.html](http://slr.colorado.edu/beginweb/speech_recognition/sonic.html)

<sup>9</sup> See <http://gate.ac.uk/>

deals with English documents and will be shortly extended to the other languages of the project, namely Dutch and German.

## **5 KNOWLEDGE FUSION, CREATIVE & CAMPAIGN DETECTION**

Creative detection, campaign detection and cross-media interlinking are addressed by several sequential steps. At the very beginning of the process different advertisement appearances are tracked. Then they are automatically compared to each other to differentiate new creatives from existing ones. This decision making process cannot be fully automated, thus to achieve reliable precision the ambiguous spots are shown to a user to decide. This process is further detailed in section 5.1.

The next step in processing pipeline is to form campaigns of creatives. This task is transformed to the problem of identity resolution e.g. defining whether a given creative belongs to one of the already formed campaigns or a new one. It is addressed by semantic comparison and clustering algorithms embedded to a common framework. The framework uses semantic knowledge representation based on MEPCO ontology. The data transaction between the two logical steps in campaign detection process, namely creative detection and campaign discovery, is made by the Knowledge Fusion module. It is responsible for the translation of a creative metadata description into ontological knowledge representational formalism. Thus it serves as an interface between the corresponding modules for creative and campaign discovery. This is further detailed in section 5.2.

### **5.1 Creative Detection**

The creative detector takes all analysis results of an investigated creative as input and generates a combined analysis result. This result indicates if the investigated creative is a new creative or if it belongs to an existing one. For existing creatives, information about the original creative is provided, and the investigated creative is marked as auto verified. The operator of the system doesn't need to inspect these results because the rate of correctly auto verified creatives should be very high. On the other hand, the operator has to manually inspect creatives which are marked as new ones from the creative detector. To facilitate this manual work, the creative detector provides a list of existing creatives which are most similar to the investigated creative, as well as the name of the company and product the creative most likely belongs to. However, the quality of the creative detector results largely depends on the results of the specific analysis modules because it doesn't make any analysis on its own.

To generate results, the creative detector combines the analysis results of the specific analysis modules in the following four steps: (1) First relevant information of all

analysis results belonging to an investigated creative is extracted. In this context the only relevant information is the similarity of an investigated creative to an existing creative, to a product or to a company. (2) Next the creative detector weighs this information according to the analysis modules they were extracted from. The separated precision and recall values of the analysis modules are used as weighting criteria. (3) Then the information is fused on an individual level. If different analysis modules have found a similarity to the same existing creative (product or company) then the probability raises that the investigated creative belongs to this existing creative. (4) For the last step, a fusion on different levels is performed. For instance, if one analysis module finds out that the input spot is very similar to an existing creative, and another analysis module recognizes the corresponding product name in this creative, then the probability raises this existing creative and the company.

### **5.2 Knowledge Fusion & Campaign Detection**

The Knowledge Fusion is a bridge between the Creative Detection and the Campaign Detection modules. Its role is to convert the information stored in the MMS repository about a new Creative into a representation suitable for the Campaign Detection component, using a RDF (Resource Description Format) representation of a Creative based on the MEPCO ontology [1]. The campaign detection module is based on the Identity Resolution Framework (IdRF), which is intended to be a general framework for identity resolution with respect to an ontology. The framework contains a Semantic Description Compatibility Engine (SDCE) used for computing a similarity measure between instances based on a collection of rules. An initial step in the IdRF restricts the comparisons to a subset of entities stored in the knowledge base, the remaining instances are then compared to the new Creative passed as input. The similarity scores are subsequently used for attaching the Creative to an existing Campaign or generate a new one. Information about Campaigns is stored in the OMS and will be displayed and queried by the Delivery System of the MediaCampaign prototype, which will relate it to the data stored in the other repositories (EMS, MMS). This will allow it to display media spots associated to a Creative and see how they are related to others inside a Campaign. The web interfaces of the KIM platform are currently used for debugging and displaying the content of the OMS.

## **6 SUMMARY AND CONCLUSIONS**

We introduced the first prototype of a complex system for new creative and campaign discovery within the media Press, Internet and TV. The MediaCampaign prototype utilizes four modalities (image, text, video and audio) and covers three languages (English, Dutch and German). The system is mainly based on a multimodal semantic analysis subsystem for new creative detection and on an identity

resolution framework for campaign discovery and tracking further supported by the MEPCO ontology. Furthermore many additional components are involved mainly for user interaction. All components are integrated via a sophisticated architecture based on flexible web services. From the users' point of view a press/Internet and TV workflow can be distinguished. In connection with image/video fingerprinting, logo recognition and jingle recognition we could already prove by initial tests on a limited dataset, that the chosen methodologies are promising and accurate in terms of precision and recall.

MEPCO ontology in connection with the chosen identity resolution method turned out to be sufficient to reveal very first evaluation results in terms of campaign discovery. Out of currently 6366 creatives contained in the MediaCampaign knowledge base we could find 536 distinct campaigns.

## 7 FUTURE WORK

The currently set-up first prototype of MediaCampaign supports within its text and speech analysis components only the language English. In the next project year we will extend the text analysis modules and the ASR subsystem to support also the languages Dutch and German. Last but not least we will extensively evaluate the first MediaCampaign prototype. Especially in terms of precision and recall in connection with the semantic analysis subsystem and the campaign discovery. Also we already scheduled first end user evaluations especially with regards to the press workflow subsystem.

## 8 ACKNOWLEDGEMENTS

The R&D work carried out for the MediaCampaign project is partially funded under the 6<sup>th</sup> Framework Programme of the European Union within the strategic objective "Semantic-based knowledge and content systems" of the IST Workprogramme 2004 (IST FP6-027413).

## 9 REFERENCES

- [1] B. Popov, P. Todorova.: Media Presence and Campaign Ontology (MEPCO), public available for download including description from: <URL: <http://www.media-campaign.eu/>>
- [2] H. Rehatschek.: DIRECT-INFO: Media monitoring and multimodal analysis for time critical decisions. Proc. of 5<sup>th</sup> WIAMIS conf. 2004, ISBN-972-98115-7-1, Lisbon, (2004).
- [3] H. Rehatschek, N. Diakopoulos, G. Kienast, V. Hahn, T. Declerk.: DIRECT-INFO: A Distributed Multimodal Analysis System for Media Monitoring Applications. Proc. of EWIMT, ISBN 0-902-23810-8, London, (2004), pp. 367 – 374.
- [4] W. Haas, H. Mayer, G. Thallinger.: "Real Time Monitoring of Radio and TV Broadcasts". CBMI 2003 - Third International Workshop on Content-Based Multimedia Indexing, September 22-24, 2003, IRISA, Rennes, France, (2003).
- [5] Covell, M.; Baluja, S.; Fink, M.: Advertisement Detection and Replacement using Acoustic and Visual Repetition Multimedia. Signal Processing, 2006 IEEE 8th Workshop on, Vol., Iss., (2006), 461-466.
- [6] Xian-Sheng, H.; Lie, L., Hong-Jiang Zh. : Robust learning-based TV commercial detection Multimedia and Expo, 2005. ICME 2005. IEEE Int. Conference, Vol., Iss. (2005).
- [7] Weiss, J.: Genre classification: Semantic Interpretation of Video. Diploma thesis, Graz University of Technology, faculty of computer science, Institute for Computer Graphics and Vision, 2005.
- [8] H. Saggion, H. Cunningham, K. Bontcheva, D. Maynard, O. Hamza, Y. Wilks. Multimedia Indexing through Multisource and Multilingual Information Extraction; the MUMIS project. Data and Knowledge Engineering, 2003.
- [9] M. Dowman, V. Tablan, H. Cunningham and B. Popov. Content Augmentation for Mixed-Mode News Broadcasts. 3rd European Conf. on iTV: User Centred ITV Systems, Programs and Applications. Aalborg Univ., Denmark, 2005.
- [10] H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. Proc. of the 40th Anniv. Meeting of the Association for Computational Linguistics (ACL'02). Philadelphia, July 2002.
- [11] Scheirer, E.; Slaney M.: Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator. ICASSP, 1997.
- [12] Ajmera, J; McCowan, I; Bourlard, H: Speech/music segmentation using entropy and dynamism features in a HMM classification framework. Speech Communication. Vol. 40, no. 3, pp. 351-363. May 2003
- [13] Pinquier, J.; Senac, C.; André-Obrecht, R.: Speech and music classification in audio documents, ICASSP, 2002.
- [14] Cano, P., Battle, E., Kalker, T., and Haitsma, J. 2005. A Review of Audio Fingerprinting. Journal of VLSI Signal Processing Systems 41, Nov. 2005, pp. 271-284.
- [15] Ellis, D. and Poliner, G., Identifying 'Cover Songs' with Chroma Features and Dynamic Programming Beat Tracking Proc. Int. Conf. on Acous., Speech, & Sig. Proc. ICASSP-07, Hawai'i, April 2007, pp. 1429-1432.
- [16] J.S.Garofolo, C.G.P.Auzanne, and E.M. Voorhees. The TRECS DR Track: A Success Story. In: Eighth Text Retrieval Conference, pp. 107-129, Washington, 2000.
- [17] de Jong, F.M.G. and Ordelman, R.J.F. and Huijbregts, M.A.H. (2006) Automated speech and audio analysis for semantic access to multimedia. In: Proc. of the 1<sup>st</sup> Int. Conf. SAMT 2006, Athens, Greece. pp. 226-240.
- [18] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada. MPEG-7 color and texture descriptors. IEEE Trans. Circuits and Systems for Video Technology, 11:703–715, June 2001.
- [19] David G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2):91–110, 2004.
- [20] Paar, G. and Pölzleitner, W., Robust Disparity Estimation in Terrain Modeling for Spacecraft Navigation, in Proc. 11th ICPR, International Association for Pattern Recognition, 1992.