

MediaCampaign Public Annual Report 2006



<http://www.media-campaign.eu>

MediaCampaign's scope is on discovering, inter-relating and navigating cross-media campaign knowledge. A media campaign is a collection of 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 media campaigns on television, the Internet and in the press.

A multimodality technology combining audio, video and text analysis for the detection and tracking of campaigns is being developed by the project. A prototype will be implemented for the application domains of information technology and the automotive industry.

MediaCampaign began in April 2006 and will run until September 2008.

Summary of Activities

MediaCampaign's first year has been used to establish the user requirements and define a use-case, to define a first test data set, to set-up a first ontology for campaign modelling and for the definition of the overall system architecture. A prototype of the system will be used internally by the exploiting partner. With connection to test data a set containing advertisements in the press, the internet and television has been issued and covers the automotive and information technology domains.

MediaCampaign aims at using a semantic analysis of advertisements in order to discover and track media campaigns. This semantic information is modelled and represented using the Media Presence and Campaign Ontology (MEPCO). A first version of the MEPCO ontology has been defined and published. A community portal is available which allows raising and tracking issues regarding the development of the ontology.

One of the main achievements of the first year was the definition of the initial system architecture with the high level building blocks, identification of the interfaces between these modules and of the technical workflow. A number of rapid prototyping tools have been defined in order to try out at an early stage the performances of the different analysis modules using the test data set provided by the exploiting partner. Improved versions of the subsystems and components and their integration in the MediaCampaign prototype are the main targets for the second year.

Important work area

User requirements

The user requirements for the MediaCampaign project have been specified by Nielsen Media Research UK (NMR-UK). Nielsen Media Research is the global leader in multinational media research and analysis.

The scope of MediaCampaign goes beyond the state of the art by providing not only information about which creatives are already known in the database or not, but by relating the creatives over different languages and modalities. It is also expected that the identification of existing creatives will be improved by the technology developed by MediaCampaign, notably by facilitating the manual coding of metadata.

Nielsen Media Research has provided a test dataset of adverts which will be used in order to test and tune the prototype. The test dataset pertains to the sectors of IT and the automotive industry. The spots are in English, Dutch and German and have been taken from 3 TV channels, 16 newspapers, 13 magazines and 5 internet websites.

The test dataset is split in two subsets:

- newly collected data: one month (for TV, Press, Internet)
- 1 year's worth of creatives for the selected country and sectors, with the detail per media type, sector and country.

The prototype released by the consortium will be used and evaluated by Nielsen Media Research UK.

MEPCO Ontology

The identification and tracking of a media campaign relies on prior identification, recognition and interrelating of the creatives that made up the campaign. This means that various analyses of the different modalities are to be performed in order to extract features related to campaigns of all types. Four modalities will be analysed and cross-linked – image, text, audio and video. The use of an ontology for the identification and for the tracking of adverts and campaigns is an innovative approach within the MediaCampaign project.

An extensive study of related standards has been carried out. MEPCO is an extension of the PROTON ontology, which has good coverage of concepts, shared across domains and provides a means for a wide range of tasks such as semantic annotation, indexing and retrieval of documents.

MEPCO has a general cross-media level and specialised extension for different media. It deals with concepts like Creative, Spot, Campaign, Ad, Commercial, Media etc. and models their attributes and properties. The ontology is encoded in OWL Lite. This sublanguage of OWL is backward-compatible with RDF(S).

A community portal for MEPCO open for the general public is available on <http://ontofile.sirma.bg:8080/MEPCO> and is used to help the maintenance and further development of the ontology.

Architecture

The specification of the System architecture is one of the major achievements of the first year of MediaCampaign. The main focus was put on the definition of high-level building blocks, the interfaces between them, the description of the overall technical workflow, the information flow and the data flow. The technical workflow was derived directly from the User Requirements, and Use Cases defined by NMR.

The technical workflow consists of the four major building blocks:

- (1) **Media Acquisition:** The acquisition will be carried out on a (potentially) distributed system, and is well suited for the project needs. The component centralises the acquisition of all media essences and decouples the system from the different media sources. The acquisition component will accept for the media TV and press digital file input (single creatives), for Internet an appropriate grabber will be provided.
- (2) **Media Analysis:** 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.
- (3) **Knowledge Fusion and Campaign Discovery:** Once all the modalities of an item (a spot) have been analysed (sound – image – video - text), the creative is compared with 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.
- (4) **Delivery System:** The delivery component will present the overall results of the system. The delivery system will focus on campaigns and will provide to the operator the results of the cross media interlinking analysis. The aim of the delivery system is to browse, check and validate data regarding campaigns in order to present them to final customers. Customer delivery could be performed by an external system.

User Involvement, Promotion and Awareness

The MediaCampaign consortium includes Nielsen Media Research UK, which is a part of the VNU group. Nielsen Media Research is the global leader in multinational media research and analysis. The company is active in 40 markets, offering television and radio audience measurement, advertising information services, print readership and customised research services. NMR has specified

the user requirements for the MediaCampaign system and will be the main user of the prototype delivered by the project. This ensures that the technology is fit for purpose and that MediaCampaign will lead to new business cases in media monitoring and analysis, and positively impact the European media monitoring sector.

The MediaCampaign consortium took part in a tutorial held in the framework of the European Semantic Web Conference (11-14 June 2006). The tutorial was jointly organised by the University of Sheffield and Joanneum Research, and presented the latest results from several European projects on multimedia and semantically-enabled knowledge technologies, such as Prestospace, NM2 and Sekt), which made this tutorial a significant concertation activity amongst these research projects.

Events in connection with MediaCampaign

The activities of the MediaCampaign project have been presented in the following papers and tutorials:

Poster and overview paper at EWIMT2005
(<http://www.acemedia.org/ewimt2005/>)

Paper presented at the workshop “*Crossing media for improved information access*” held in the framework of the 5th Language Resources and Evaluation Conference - LREC 2006, on 22-28 May 2006 in Genoa, Italy (<http://www.lrec-conf.org/lrec2006>)

Tutorial “*What you Mean is What you Watch: Multimedia and the Semantic Web*” during the European Semantic Web Conference (<http://www.eswc2006.org>)

Tutorial “*Human Language Technology (HLT) for Knowledge Acquisition for the Semantic Web*” 15th International Conference on Knowledge Engineering and Knowledge Management (EKAW 06 - <http://ekaw.vse.cz/>)

Presentation at ESCW 2006 by Michael Hausenblas
(<http://www.eswc2006.org/tutorials.html>)

Contribution to the brochure for FFG (<http://www.ffg.at/>)

Accepted paper for SAMT2006 (December 2006): “*Automated speech and audio analysis for semantic access to multimedia*”

Future Work

During the next period, the first integrated prototype will be completed and evaluated based on the test plan. Feedback will be provided to system developers by the exploiting partner NMR. The system will be tested on the domains of information technologies and the automotive industry, and in three languages, English, German and Dutch. This will include a first version of the delivery system, navigation and reporting components which will provide end-user functionalities for access to, presentation and exploitation of semantic analysis results.

The individual components for advertisement analysis of modalities audio, video and text will be evaluated using the test data set provided by the user partner and improved versions of them will be released. A first version of the Campaign Discovery module will also be made available and will be used to find relations between creatives over different languages and media inside a campaign.

The MEPCO ontology will be improved following the first prototype and will model the additional information provided by the analysis modules.

Consortium Partners

Joanneum Research (A). (coordinator)
<http://www.joanneum.at/iis>

Nielsen Media Research (UK)
<http://www.nielsenmedia.co.uk>

University of Sheffield (UK)
<http://www.sheffield.ac.uk>

Sirma Group Corp (BG)
<http://www.ontotext.com>

HS-Art Digital Service GmbH (A)
<http://www.hs-art.com>

Softeco Sismat SpA (I)
<http://www.softeco.it>

University of Twente (NL)
<http://hmi.ewi.utwente.nl>

Netherlands Organisation for Applied Scientific
Research (NL)
<http://www.tno.nl>