



12-monthly Activity Report Period 01

Public Executive Summary

MediaCampaign Identifier: MC-T12-JRS-04-PublishableSummary_P01.doc

Author(s) and company: Herwig Rehatschek (JRS), Helen Hasenauer (JRS),
Herwig Zeiner (JRS), Mark Guinney (NMR-UK), Julien
Nioche (USFD), Borislav Popov (ONTO), Walter
Plaschzug (HSA), Marco Masetti (SOF), Roeland
Ordelman (UT), David van Leeuwen (TNO)

Work package/task: WP 1 / T1.1

Document status: Final

Confidentiality: public

Keywords: project highlights, project progress

Abstract: This document contains the public executive summary
of the periodic activity report for the period 04/2006 (PM
01) – 03/2007 (PM 12) of MediaCampaign project.

DOCUMENT HISTORY

| Version | Date | Reason of change |
|----------------|-------------|-------------------------|
| 1 | 2007-05-07 | document created |

Table of Contents

| | |
|--|----------|
| 1 Publishable Executive Summary | 1 |
| 1.1 Major Achievements within the first project year | 2 |
| 1.2 Next steps | 3 |
| 2 Glossary | 4 |

1 Publishable Executive Summary

MediaCampaign targets Action Line IST-2004-2.4.7 "Semantic-based knowledge and content systems" of the IST work programme 2004 (IST FP6-027413). 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.

Within the prototypes developed in MediaCampaign project the focus will be laid on a concrete example for a media campaign: advertisement campaigns (e.g. a new car model is introduced into the market). An advertisement campaign consists of a number of single advertisements (creatives) on the same product issued in different media and/or countries. Up to four modalities (image, video, text, audio), within the targeted media (press, TV, Internet) are analysed and combined by the project. Three languages (German, Dutch and English) are covered. A prototype will be implemented for the application domains of information technology and the automotive industry.

This will lead to new business cases in media monitoring and analysis, and positively impact the European media monitoring sector. In support of this goal, the project addresses the following objectives:

- Modelling the knowledge present in campaigns and their related business and social processes;
- Analysis of tele-visual and image-based media streams and parallel focussed web analysis to extract features related to campaigns of all types;
- Fusion of the diverse evidence from the different media and different analysis methods;
- Delivery of the campaign data in an end-user pilot prototype.
- Fusion of the diverse evidence from the different media and different analysis methods, with innovative methods to exploit their complementarity.

For the pilot system developed within the project we will focus on a concrete example for a media campaign: advertisement campaigns e.g. a new car model is introduced into the market, however, the system will be designed in such a way that it can be extended to monitor and analyze also other campaigns such as political campaigns. Hence a major technical objective is to design the system architecture as flexible as possible by using well defined interfaces and utilizing standards for all components where possible. Scientifically MediaCampaign discovers three main innovations as given below.

Ontology for cross-media modelling of media presence and campaigns: a first version of the Media Presence and Campaign Ontology (MEPCO) was defined, allowing for modelling of media campaigns, including media presence information. MEPCO has a general cross-media level, extended with fine-grained specifications for the different media. Overall, the project will exploit the advantage of the Semantic Web technologies to model, manage, query and navigate within multi-dimensional information spaces with non-linear dimensions.

Semantic cross-market product data interlinking: Due to cultural differences and national market specifics similar or analogous products are branded and marketed differently across the countries. This is handled through semantic representation and mapping of brand and product related data (catalogues, feature-related terminology, etc.). MEPCO allows for (i) extension with domain ontologies specific to particular markets and (ii) cross-linking artefacts (e.g. products) between those.

Identification & tracking of new media campaigns in different media and countries: This is realized by (i) combined multimodal analysis of the media content (text, audio and video), (ii) usage of a semantic media store to consolidate and query the analysis results and all the related materials, (iii) detection and tracking of campaigns based on heuristic patterns and statistical analysis of "popularity" timelines and "media distance" metrics.

1.1 Major Achievements within the first project year

As a first step partner Nielsen Media Research has specified the user requirements and provided a test dataset of adverts which will be used in order to test and tune the prototypes developed within MediaCampaign project. The test dataset relates to the IT sector and the automotive industry. In parallel and synchronized with the user requirements by a WP meeting the system architecture has been specified. This document describes the overall system architecture and technical specification of the MediaCampaign system components. The main focus was on the definition of high-level building blocks, the interfaces in between them, the description of the overall technical workflow, the information flow and the data flow. Main steps in the workflow include (1) Media Acquisition, (2) Media Analysis, (3) Knowledge Fusion and Campaign Discovery, and the (4) Delivery System.

In connection with the media analysis first prototypes of all semantic analysis modules have been developed. In detail for audio analysis an audio segmentation module, a speech recognition system for E and NL and a word spotting module are available. In connection with visual analysis modules for image fingerprinting, video fingerprinting, image and video OCR and logo recognition are available. Last but not least also a text analysis was developed. The Text Analysis module finds instances of MEPCO (Media Presence and Campaign Ontology) concepts such as advertiser and product in textual documents obtained from OCR or ASR. In order to efficiently control the distributed multi-modal analysis modules a central workflow component called collection processing manager (CPM) was developed. It enables a system operator to monitor the analysis state of the MediaCampaign system.

With regards to knowledge fusion, campaign detection and tracking also a first prototype has been developed which combines the information generated by the analysis modules, converts it into concepts of the MEPCO ontology and finds relations between instances of these concepts. The New Campaign Detection and Tracking functionalities are done by the Campaign Discovery module, which is based on the Identity Framework. This module has then been designed and implemented and is used as a basis for the Campaign Discovery service, which communicates with the Knowledge Fusion module via a Web Service. This module is already functional and finds relations between Creatives within Campaigns. The information is stored in the Ontology Management Service and will be accessible from the delivery system later in the project.

A Media Presence and Campaign Ontology (MEPCO) has been developed as a domain specific extension. It allows for modelling of media presence information (e.g. references, articles, channels). MEPCO has a general cross-media level, and specialised extension for different media. It deals with concepts such as Creative, Spot or Campaign and models their attributes and properties. MEPCO provides a common conceptual framework for the other components within the project. That is for media analysis, new campaign detection and tracking components, as well as for the delivery system. The first version of the MEPCO ontology has also been defined and published on the public MediaCampaign website. A community portal for MEPCO open for the general public is available, and is used to help the maintenance and further development of the ontology. It has been designed according to the analysis of the domain and existing terminology and models. Significant effort has been focused on encompassing notions to be needed in future extensions towards more general campaign modelling. The concepts of campaign, creative, spot, their attributes and relationships were introduced. These were then evaluated by the partners and feedback resulted in fruitful adjustments and refinements. The MEPCO ontology has been also evaluated by modelling real-world instances from the media campaign domain, namely advertisements from printed media. This revealed improvements to be made with respect to actual modelling of the entities.

There was a very strong and focused activity on the preparation of scientific papers for dissemination purposes. As a consequence the project MC could achieve already a very prominent number of six scientific publications. These scientific results have been made already available on the public media campaign website, amongst them the MEPCO ontology. Further scientific results of the project will be made public available, in order to boost the research advances of the European research community. The exploitable results of MediaCampaign are mainly software, which will make it possible for the project partners to broaden their portfolio of products and initiate commercial activities.

One example for an early exploitation objective was provided by partner HS-Art. They want to use the results of MediaCampaign for their system Prima-RS (Advertisement recognition and Management System for Print Media). HS-Art has strong experience in the commercialisation of research results by developing tailored IT solutions based on all fields of digital media. By exploiting the technology developed in MediaCampaign, HS-Art expects to strengthen its position in a unique niche market.

1.2 Next steps

The first integrated MediaCampaign prototype will be finalized end of May 2007 and demonstrated in June 2007 during the EC review. All major components shall be available at this time in a first draft version and integrated via web service communication interfaces. Major components include the AdClipper, AdComparer, TVComparer, new creative detectors for Press/Internet and TV, knowledge fusion and campaign detection, delivery system and the knowledge stores EMS, MMS and OMS.

2 Glossary

Terms used within MediaCampaign project sorted alphabetically.

- API** Application Programming Interface
- Campaign** A campaign represents a number of creatives semantically belonging together. A campaign has a certain duration and can be cross country and cross media. See also "spot" and "creative"
- Creative** A Creative represents all the occurrences of a similar spot. A Creative is not cross media (e. g. TV and press) and not cross language. See also "spot"
- CPM** Collection processing manager. Central component which controls the analysis workflow via web services.
- DB** Database
- DBMS** Database Management System
- EMS** Essence management store – holds all essence of MC, i.e. TV spots, press images and meta essence such as extracted audio files.
- JAPE RULES** Extraction patterns used to identify new entities
- HTML** Hypertext Markup Language
- HTTP** Hypertext Transfer Protocol
- IDE** Integrated Development Environment
- IE** Information Engineering
- iFS** Internet File System
- ISO/ANSI** International Organization for Standardization / American National Standards Institute
- META-ESSENCE** Under this term we understand all essence and additional data which is generated by subsystems (e.g. lo-res videos, keyframes, OCR text, ...).
- MMS** Metadata Management Store – holds all data extracted by the analysis systems, i.e. low-level and mid-level features such as tagging of spots in connection with logos.
- OMG** Object Management Group
- OMS** Ontology management store – holds all semantic knowledge of MC, i.e. the MEPCO ontology and the campaign knowledge.
- OS** Operating System
- PL/SQL** Procedural Language/SQL
- Re-detection** The same creative is re-detected in a single media. E.g. the ad of Renault Megane appeared on 5 May 2005 on "TV ORF1" 5.00 pm and on 7 June 2006 on "TV ORF1" at 4.00 pm.
- Scene** A consecutive series of shots connected through transitions (hard cut, fade, panning, ...) that constitutes a logical unit of action in a video. It is defined by "bridging features" such as same visual/audio content, music/speech/noise segments or text overlays.
- SGML** Standard Generalized Markup Language
- Shot** A consecutive series of pictures representing coherent visual content, e.g. when having an interview with two persons, and the camera is changing

between the two faces (depending on who is talking), each face would be a shot.

SOAP Simple Object Access Protocol

SQL Structured Query Language

Spot An occurrence of a Creative, e.g. 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 a new or existing Creative. See also "creative".

S&R Search and Retrieval

Tracking Tracking of creatives by means of history over media and countries. When a creative is re-detected (see also "re-detection"), it is inserted into a data structure which represents the history of this creative. This enables users to query the database for appearances of this creative in the past (e.g. it had appeared on 1 April 2006 in press "The Times" in US p.5 and on 3 April 2006 on "TV Sky 1" at 19.00 in UK.

WP Work Package

UML Unified Modelling Language

URI Uniform Resource Identifier

XML Extensible Markup Language

XPath XML Path Language

XSLT XSL Transformation

Partner Acronyms:

HSA HS-Art Digital Service GmbH

JRS JOANNEUM RESEARCH Forschungsgesellschaft mbH

NMR-UK Nielsen Media Research UK

ONTO Sirma AI EAD

SOF Softeco Sismat SpA

TNO Netherlands Organisation for Applied Scientific Research

USFD University of Sheffield

UT University of Twente