



MEPCO Version 2

Deliverable 3.4

MediaCampaign Identifier: MC-D34-ONTO-04-MEPCO-Version2.doc

Author(s) and company: P. Todorova (ONTO)

Work package/task: WP3

Document status: Final

Confidentiality: Public

Keywords: Ontology, properties, semantics

Abstract: Final version of MEPCO ontology taking also metadata from the analysis modules into account

DOCUMENT HISTORY

<i>Version</i>	<i>Date</i>	<i>Reason of change</i>
1	2008-02-19	First draft (P. Todorova)
2	2008-03-17	Editing after peer review (P. Todorova)
3	2008-03-18	Final editing by JRS

Table of Contents

1	Executive Summary	2
2	Introduction	3
2.1	Purpose of this document.....	3
2.2	Scope of this document.....	3
2.3	Status of this document.....	3
2.4	Related Documents	3
3	Current state of the ontology.....	4
3.1	Brief introduction to MEPCO ontology.....	4
3.2	Brief summary of the first release of the ontology	6
3.3	Changes in the second release of MEPCO ontology.....	6
3.4	Mappings between the output metadata and MEPCO	6
3.5	Changes made in MEPCO ontology	8
3.5.1	<i>Changes in the structure of the ontology</i>	<i>8</i>
3.5.2	<i>Introducing new properties in the ontology</i>	<i>8</i>
3.5.3	<i>Renaming properties in the ontology</i>	<i>12</i>
3.5.4	<i>Removing properties and classes from the ontology.....</i>	<i>12</i>
3.5.5	<i>The new property containsFocusWord explained.....</i>	<i>13</i>
4	MEPCO Ontology – naming and versioning.....	15
4.1	Naming	15
4.1.1	<i>Obligatory elements</i>	<i>15</i>
4.1.2	<i>Optional elements.....</i>	<i>15</i>
4.2	Versioning.....	15
5	Conclusions.....	17
6	References	18
7	Glossary	19

1 Executive Summary

This deliverable describes the changes in the second release of MEPCO ontology and the rationales behind them. The objectives of introducing the latest changes were to establish a mapping to output metadata of the various analysis modules in the first prototype and to allow for an easier discovery of related creatives and thus the discovery of ongoing advertising campaigns in media. Along with the description of changes, this deliverable documents the naming and versioning conventions to be followed in the releases of MEPCO ontology and instance data. These conventions should be applied to any subsequent releases of the ontology.

The main focus of this deliverable is to represent the latest changes in MEPCO ontology. Brief explanation of the ontology is also provided. However if one expects to gain a complete understanding of the ontology it is necessary to read two other deliverables: MC-D32-ONTO-04-MEPCO-Design deliverable describes all main classes and properties in the ontology; MC-D33-ONTO-03-MEPCO-version-1 describes the changes made one year after the first release of the ontology. Together these three deliverables are the complete set which documents MEPCO ontology.

2 Introduction

2.1 Purpose of this document

The purpose of this document is to give a comprehensive overview of the latest changes in MEPCO ontology; to justify the changes and to describe their connection with the overall purpose of Media Campaign Project.

MEPCO ontology models the domain of advertisements as well as advertising campaigns. To allow for the analysis and easier discovery of related advertisements new elements were introduced into the ontology.

Chapter 3 lists and explains the changes made in the current version of the ontology

Chapter 4 supplies naming and versioning conventions to be used in the file name and in the code of the ontology and instance data.

2.2 Scope of this document

This deliverable explains which classes and properties of MEPCO ontology were changed and why. After reading the deliverable, the user will know the rationales behind the changes in the ontology and will have an impression what is the intended use of the main classes and properties.

2.3 Status of this document

This document is a working draft.

2.4 Related Documents

MC-D61-JRS-09-SystemArchitecture is a document in which the overall system architecture of the Media Campaign system is defined. It contains a description of the main functional blocks and their interrelation within the overall architecture. It also gives the reader understanding of the place MEPCO has in the system architecture.

MC-D21-NMRUK-04-UserRequirements is a document with the purpose to accurately assess end-user needs. It is dedicated to defining and characterising the operational environment and its conditions. The requirements thus defined were taken into account in the design of MEPCO ontology.

MC-T62-JRS-09-MMS_UserManual is a document that describes the MMS Web Service component, part of the overall Media Campaign System Architecture. It presents the functionalities of the web service and what concerns MEPCO refinement, it presents also the MMS metadata.

MC-D32-ONTO-04-MEPCO-Design is a document that provides a detailed overview of MEPCO ontology, together with a section on methodology used, design principles followed and also with comprehensive presentation of the hierarchy of classes and their properties.

MC-D33-ONTO-03-MEPCO-version-1 describes the refinement of MEPCO ontology and the rationales for the changes.

MC-D642-JRS-01-User Manual describes the output of all the modules in the system architecture. This output contains metadata to which MEPCO classes should be aligned.

3 Current state of the ontology

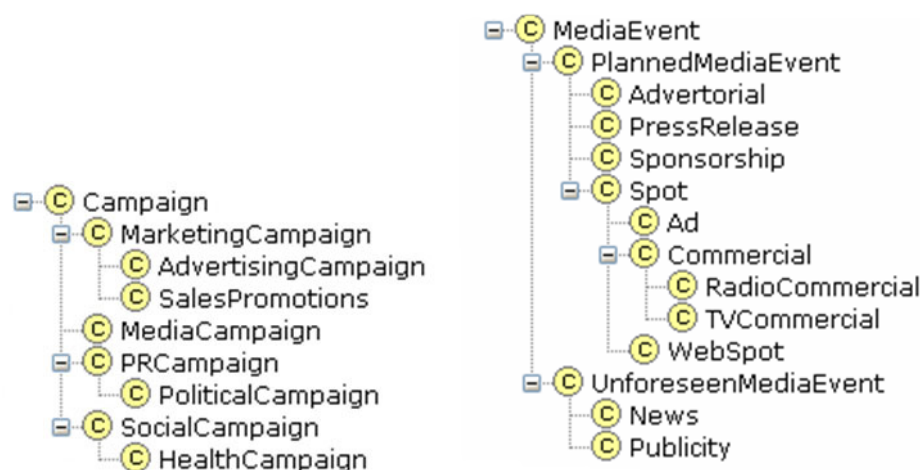
3.1 Brief introduction to MEPCO ontology

Media Campaign and MEPCO ontology

The objectives of Media Campaign Project are to discover, interlink and track new media campaigns. Media campaigns are campaigns which use mainstream media to address their target group. It is possible that a given campaign runs not only on different media at the same time but also in different countries. The main goal of Media Campaign project is to detect and track automatically such media campaigns. To do so, MEPCO ontology is introduced as a means to model the domain of advertisements and the domain of campaigns.

The classes and properties of MEPCO model significant elements and features of advertisements appearing in the three main types of media (TV, press and the Internet). In this way the analysis and the discovery of related advertisements, part of an ongoing advertising campaign is enabled. In addition, it is possible to interlink and track media campaigns running in different media and in across countries. The main classes in the ontology are Spot, Creative and Campaign as shown in Figure 1. The main classes in MEPCO ontology: Campaign and Spot.

Figure 1 The main classes in MEPCO ontology: Campaign and Spot



The hierarchy of the class Campaign is such that it models all campaigns not just the media campaigns which are the main focus of Media Campaign project and thus the focus of the ontology. The modelling allows for reasoning about campaigns in general and in a sense is complete. Media campaign is a special subclass of Campaign which allows to state that a campaign runs on media. There is a special property of a campaign – *includesCampaign*, which allows to state that a given campaign has another campaign as a part.

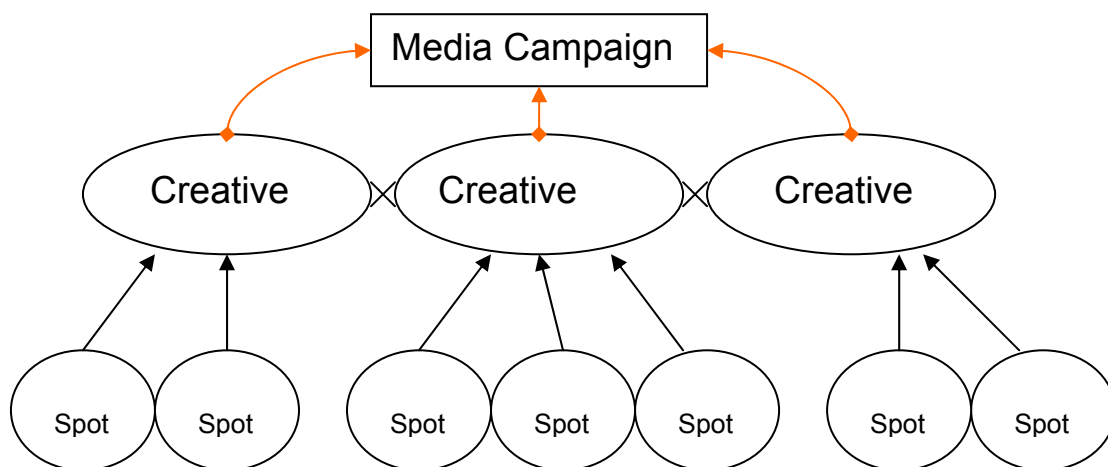
Spots are the advertisements which appear in media. They are planned, in the sense that the organization which benefits from advertising itself or its products, brand, image, etc. planned the publishing of the advertisements, it is consent with this publishing, and it is aware of the effects of the publishing. In contrast with this, the same organization or its products might be the topic of news which also affects the image of the organization. To distinguish between these two possibilities, distinct classes for planned and unforeseen media event were introduced. For the sake of completeness advertisements-like events were categorized. However, the main focus of Media Campaign project is on spots.

Creative is another central class in the ontology. It serves to model the business of the companies which monitor the media expenditures and ongoing advertising campaigns. A creative serves to summarize the important features of many similar concrete appearances of the same advertisement. It is then unnecessary to keep records of any single appearance of the same advertisement or of

advertisements which differ only slightly from each other, say in font, in size, or in the appearance of one visual element less or more.

The relationship between Spot, Creative and Campaign is presented on Figure 2 **The relationship between Spot, Creative, and Campaign** Each creative is a set of common features shared by spots, e.g. image, text of the advertisement, company name or logo, brand, etc. A creative is an abstract description of the spots and thus the arrows pointing from spots to creatives signify that some spot features are abstracted. The relationship between a media campaign and creatives is as between the whole and its parts : a media campaign consists of creatives which are semantically related to each other. To emphasize the part-Of relation between a campaign and a creative, we used an orange arrow.

Figure 2 The relationship between Spot, Creative, and Campaign



A Campaign that runs on media could be tracked by monitoring the appearing creatives and determining which of them are semantically related. The decision about relatedness of creatives is based on their semantics because their other features are not as unique and depend on the type of the media. For example, the size of an advertisement depends on the size of the media, the price paid for the publishing of the advertisements, to name but a few. One and the same advertisement may appear in several newspapers and magazines and differ slightly in terms of size. However, the logo, the text, and the image will be the same and therefore give enough clues to state the sameness of the different appearances of an advertisement.

The creatives serve to summarize the important features of many similar spots, or concrete appearances of the same advertisement. By keeping records of the common features of similar advertisements it is easier to monitor campaigns. The database of the creatives is simpler in terms of number of items and it is then sufficient only to compare a given new advertisements with a restricted set of records in the database. The actual comparison is left to the other components in the system architecture while the ontology should provide enough means for the analysis of the advertisements and creatives and to enable and ease the comparison.

MEPCO ontology has many properties which describe the relevant features of spots, as we call the concrete appearances of an advertisement. We could distinguish three types of spot features and accordingly three types of properties in the ontology. There are properties used for the semantic analysis and useful to determine the similarity between advertisements, respectively a spot and a creative. These properties also serve to track ongoing campaigns. Then there are properties which serve for the metric analysis of advertisements like placement, size, and position which serve to estimate the expenditures per publication and for the entire campaign. Finally there are properties which serve to keep records of the advertisements in a database. MEPCO ontology has a twofold use – it models the domain of advertisements and advertisement campaign and it is useful to keep a

database of them. The first and the second release of the ontology differ mainly in terms of the number and type of properties. The hierarchy of the classes is stable and changed slightly. More explanation about the concrete changes could be found below.

3.2 Brief summary of the first release of the ontology

First release of the MEPCO ontology is not as rich as the second release is. However it contained the basic classes and properties necessary for the analysis of advertisements, creatives, and campaigns. Below we present many illustrations as a means of comparison between version of MEPCO released in 2006 and the last version. Of the three main classes, only Spot had sufficient properties in the first release of the ontology. These properties stated the placement, position, and the size of spots and are useful for the estimation of the media expenditures for its publishing. The class Spot also had few properties useful for the semantic analysis like title, subtitle, logo and slogan. The class Creative and the class Campaign lacked many of the useful properties for the semantic analysis. In the second release of the ontology such properties were introduced.

3.3 Changes in the second release of MEPCO ontology

The changes in current version of MEPCO are of two types: syntactical, which aim at providing a neat version of the ontology and semantic, which change the modelling by adding new classes and/or properties.

The current version of the ontology is improved in terms of syntactical neatness of the ontology. The semantic extension means addition of new properties and few classes to the ontology. Main changes of the second release are described in [MV]. This deliverable concerns only the latest changes made in order to map MEPCO ontology to the output metadata of the first prototype as it was decided on the last project meeting. The output metadata is described in MC-D642-JRS-01-User Manual [UM]. It relates to the output of all analysis modules in the prototype. The necessity to map the output metadata to MEPCO classes and properties serves as a justification for the latest semantic changes in MEPCO.

3.4 Mappings between the output metadata and MEPCO

It follows a presentation of the mappings between the output metadata of the first prototype and the classes and properties in MEPCO ontology. The output metadata was taken from MC-D642-JRS-01-User Manual. First we tried to establish a mapping with existing MEPCO classes and properties and whenever there was none, we outlined changes to be made in the ontology. In this, we followed the methodology for introducing changes in MEPCO as it was described in MC-D33-ONTO-03-MEPCO-version-1 [MV].

Table 1 **Mappings between output metadata in first prototype and classes and properties in MEPCO** shows the mappings between the output metadata and MEPCO.

Table 1 Mappings between output metadata in first prototype and classes and properties in MEPCO

Output metadata	MEPCO classes (c) and properties (p)
text	Property, already modelled for Spot, to be modelled for Creative and Campaign
language	Property, already modelled
matchValue	Candidate property <i>hasImageSimilarity /VideoSimilarity</i> for Spot
logoName	Property, already modelled for Spot, Creative, and Campaign
music	Property already modelled for some Spots (commercials only); to be modelled for Creative and Campaign
musicID**	See the considerations below
confidence (attribute of Music)	To be modelled if there is such attribute in the metadata
speech	To be modelled if there is such attribute in the metadata

likelihood (attribute of speech)	To be modelled if there is such attribute in the metadata
speechId**	See the considerations below
jingle	Property already modelled for Spots to be modelled for Creative and Campaign
Title (attribute of jingle)	See the considerations below

Below we provide explanation of the output metadata and list the changes in MEPCO classes and properties following the order in Table 1 **Mappings between output metadata in first prototype and classes and properties in MEPCO**.

- **filter** – the User Manual for the prototype [UM] defines that filter is an attribute which “each analysis module has the possibility to use this field to categorise the analysis result in an appropriate way; e.g. the Fingerprint modules might use this field to mark results where an identical creative was found“. For now, it is not used. After discussion with the partners, the exact meaning and use of this attribute will be defined and it will be decided if it has to be modelled in the ontology.
- **text** – until now text was a property, modelled only for the class *Spot*. In the current version of the ontology, similar property was added for the class *Creative* as well. Text allows for the search which would establish similarity among creatives and this justifies its introduction in the ontology.
- **fontSize** is an attribute of text in the output metadata. It is an open issue if we could use it to define the similarity among creatives. From one side, part of the text in the advertisement will appear in larger font size but this is a clue necessary to define which part of the text are the title and the subtitle of the advertisement and which part is a plain text. However it is not enough to know the font size to establish the similarity between creatives because some magazines may type part of the text of different creatives using the same font size, say that part which is not title or logo may appear always in 12pt. For now it is an open issue to be discussed with the partners if we should model this property in the ontology
- **language** – this attribute specifies in what language the text of the advertisement is in. This is an important attribute which would allow the discovery of campaigns across countries. It was already modelled in the ontology for the classes *Spot*, *Creative* and *Campaigns*.
- **matchValue** – The User manual [UM] says that this attribute: “indicates the similarity of the result image to the query image; 1 means that the images are identically (exact matches), 0 means that no similarity is given“. In the current version of the ontology we introduced a new property for the class *Spot* which defines the similarity of a spot to an image. This property could be used to pick up the closest matches to a given image. It could be used to establish similarity among creatives as well. However there are some difficulties in modelling such a property for creatives and in current version of the ontology we abstain from that modelling.
- **logoName** – MEPCO ontology already has properties for the class *Spot*, *Creative* and *Campaigns* which model the logo. We could use the logo name as a value of these properties. The properties pointing to logo are very important for the discovery of related creatives and thus of ongoing campaigns.
- **music** – was already modelled for the class *Commercial* (those spots which are broadcasted on TV and on radio). In future we will decide if such property is to be modelled for Web Spots as well. In the current version of the ontology, similar properties were introduced for the classes *Creative* and *MediaCampaign* because like the logo, music might be a clue as to the relatedness of creatives within a given campaign.
- **musicID** could be modelled as a value of the properties which model music

- **confidence** is an attribute of music; it shows the likelihood that a given piece is music and not something else. It will be discussed further with the partners if we should model this attribute in the ontology
- **speech** is used internally by some analysis modules. There is no need to be modelled it in the ontology
- **jingle** – there was a property in the ontology which models jingle for the class *Spot*. In the current version of the ontology, similar property was introduced for the class *Creative* and *Campaign*. Like music and especially like the slogan, a jingle is a clue as to the relatedness of creatives in a given campaign.
- **title** – the jingle title could be used as value of the property jinglge

3.5 Changes made in MEPCO ontology

This section summarizes the last changes made in MEPCO so that a mapping with the output metadata of the first prototype is established and the discovery of related creatives within a given campaign is enabled.

3.5.1 Changes in the structure of the ontology

Because the first release of the ontology was matured in terms of structure and modelling, there was no necessity to introduce new classes or change the hierarchy of the existing ones. The feedback of the project partners ensured us that MEPCO classes and properties are a good modelling of the real world situation in the advertising monitoring industry. MEPCO modelling was already tested in the environment of knowledge fusion and campaign discovery and it proved to be highly useful and reflecting the special needs of the KFCD component. As we already mentioned in different parts of this deliverable, most MEPCO properties allow for a sufficient, though not complete, analysis of the advertisements and ongoing campaigns. There are properties which specify placement, position, size and the media, including program in case of broadcasted advertisements and issue in case of printed ones. These properties are sufficient to make estimates as to the price paid and thus estimates as to the media expenditures per advertisement and per campaign. There are properties in the ontology which specify semantic features of the advertisements like text, focus words, logo, slogan, etc. These properties allow to determine which advertisements are similar and to what extent and are useful in tracking ongoing campaigns. The logo is important to decide what Brand is advertised, what company is the advertiser, i.e. benefits from spreading the message to increase the awareness of the consumers towards its products, and also to estimate roughly the Brand expenditures. The slogan is repeated throughout the campaign like the jingle and serves to track ongoing campaigns. The property specifying focus words is explained below.

3.5.2 Introducing new properties in the ontology

As it was explained in the beginning of this chapter, the latest changes in the ontology were aimed at modelling the output metadata of the various analysis modules of the first prototype. We already listed that part of the metadata to which mapping should be established and explained the changes in the ontology. In what follows we will give a comprehensive summary of the changes and will provide illustrations to compare the first and second release of MEPCO.

The properties of the class *Spot* in the first release of the ontology were of two types: which describe the semantically significant elements of the spot see Figure 3 **The properties of the class Spot in the first release of the ontology**(in squares) and which describe media-specific features of the spot (in ellipses). Since the class *Spot* has differentiations in accordance to the types of the media, e.g. subclass *Ad* for advertisements in press; commercials for advertisements broadcasted on TV or radio, and *WebSpot* for advertisements in Internet, for each of these classes specific property specifying the placement, size and position is introduced. For more details refer to the detailed description of the ontology in [MD].

In the last version of the ontology, the following properties were introduced for the class Spot:

- i. *castsCelebrity* (for Commercials only)
- ii. *hasImageSimilarity*
- iii. *containsFocusWord*
- iv. *containsLexicalVector*

Of these properties the first renames already existing property in the previous release. It was named *hasArtist* which we think is not as comprehensive as *castsCelebrity*. Second property of the above list is introduced for the sake of a mapping with output metadata and the rationales for the introduction of the last two properties is explained in section 3.5.5

Figure 3 The properties of the class Spot in the first release of the ontology

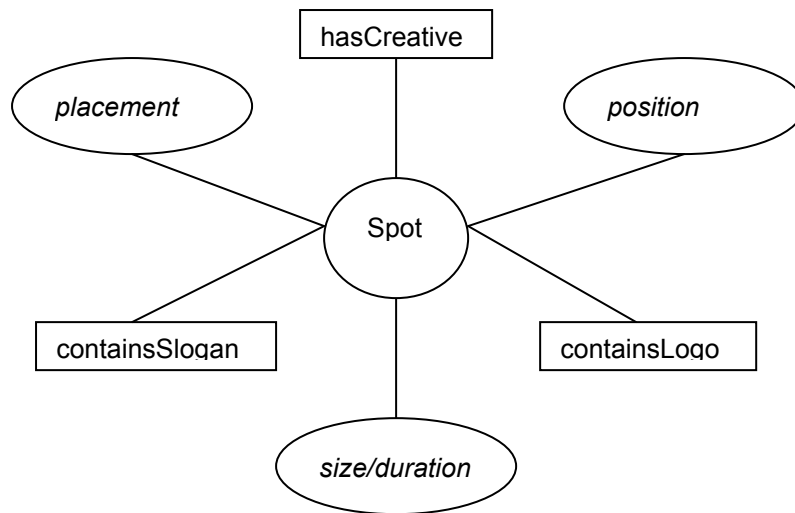
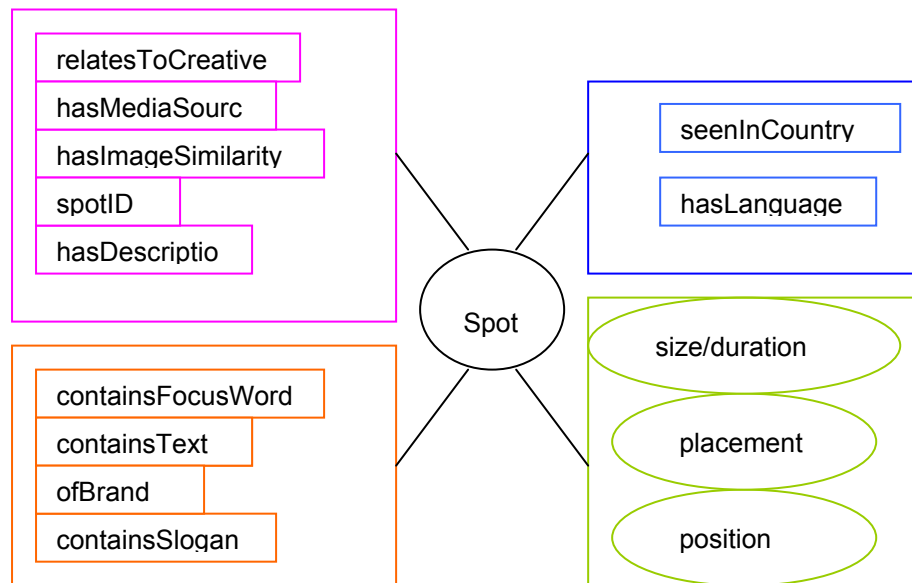


Figure 4 The properties of the class Spot in the second release of the ontology



In the second release of the ontology the class *Spot* has many more properties. It has specific properties to describe each spot as a part of/record in a database (the ones in the upper left box). It has properties which serve to track the spot across borders of the countries (the ones upper right box). It has properties which describe the semantics of the spot (the ones in bottom left box). Of these properties most were introduced during the past year.

As one goes through the illustrations, one sees that the second release of the ontology is richer. The notes above apply similarly to the properties of the class *Creative* and the class *Campaign*. In the last version of the ontology, the following properties were introduced for the class *Creative*:

- i. *textOfCreative*
- ii. *musicOfCreative*
- iii. *celebrityInCreative*
- iv. *focusWordOfCreative*
- v. *jingleOfCreative*
- vi. *lexicalVectorOfCreative*

The rationale behind the introduction of these properties is to enhance the ontology and to allow for determining the semantic relatedness of creatives.

Figure 5 The properties of the class *Creative* in the first release of the ontology

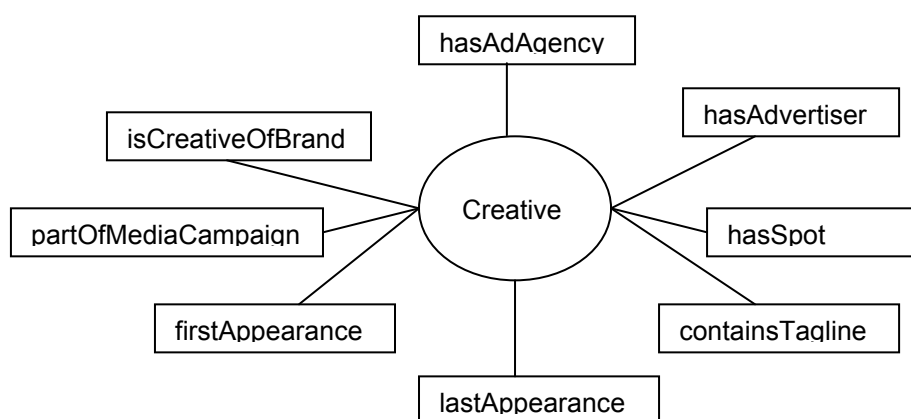
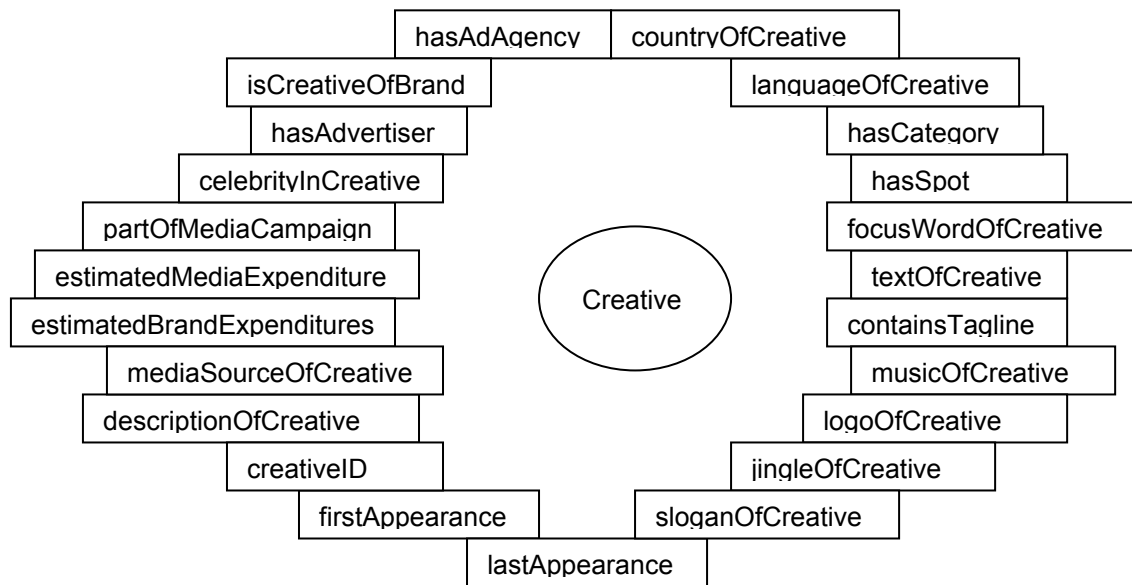
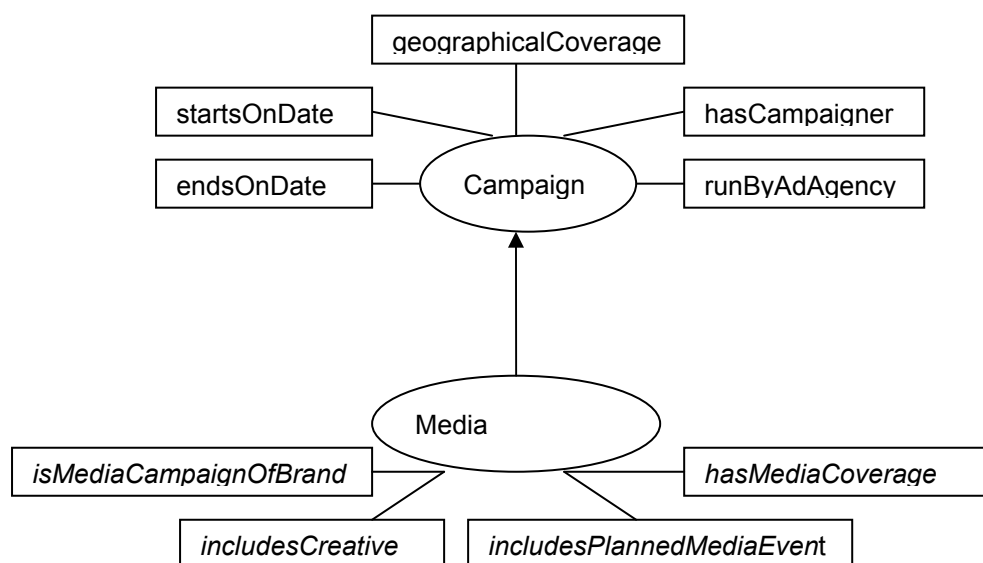
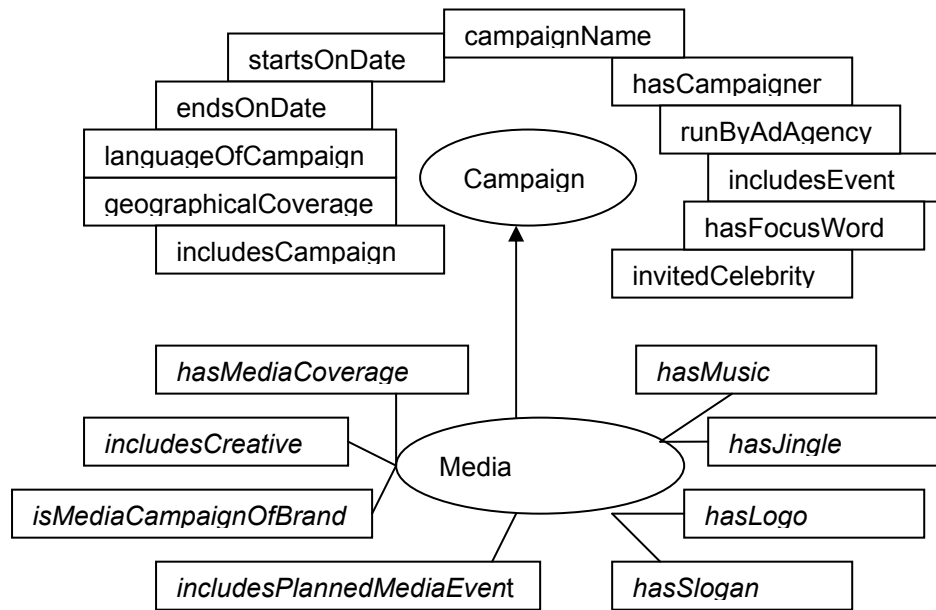


Figure 6 The properties of the class Creative in the second release of the ontology

Similarly, the properties introduced for the class Campaign aim at enabling the tracking of ongoing campaigns by means of semantics. The new properties are as follows:

- i. *hasMusic*
- ii. *invitedCelebrity*
- iii. *hasFocusWord*
- iv. *hasJingle (media campaigns only)*
- i. *hasLexicalVector*

Figure 7 The properties of the class Campaign in the first release of the ontology**Figure 8** The properties of the class Campaign in the second release of the ontology



The new properties allow for the analysis of the semantics of creatives, ease the determination of the relatedness of creatives and give enough clues to track ongoing campaigns. However MEPCO is not a complete modeling of the domain of advertisements and of campaigns and future additions which will enrich the hierarchy of the classes and properties are not excluded. In its current state it is rich enough to enable a sufficient analysis of advertisements and of campaigns, to aid keeping records in databases and to allow for determining which creative belongs to which campaign.

3.5.3 Renaming properties in the ontology

Along with the above mentioned changes which led to enhancement of the ontology or served to establish mappings to the output metadata of the analysis modules in the first prototype, some slight improvements were made in the ontology in terms of changes in the names of few properties. The changes made the names more comprehensive.

Table 2. Changes in the names of MEPCO properties

Old Name	New name
<i>brandExpenditures</i>	<i>estimatedBrandExpenditures</i>
<i>mediaExpenditures</i>	<i>estimatedMediaExpenditures</i>
<i>hasCreative</i>	<i>relatesToCreative</i>
<i>has Artist</i>	<i>castsCelebrity</i>

It is obvious that *brandExpenditures* is somewhat misleading and suggest that one knows the exact value of the expenditures, while *estimatedBrandExpenditures* are closer to the real-world situation in the advertisement monitoring industry where there are rough guesses as to the amount of money paid.

3.5.4 Removing properties and classes from the ontology

There were no properties which were removed from the ontology. However there were a single class *MediaSource* which was removed because it was a repetition of the semantics of another class – *Media*. To retain the desired semantics of *MediaSource*, a new property was introduced “*mediaSourceOfCreative*” with domain *Creative* and range *Media*. Similarly such property was introduced also for *Spot* – *hasMediaSource*; again with range *Media*. The latter renamed the property *hasMediaSource* from the previous version of the ontology. The semantics is the same but the modeling is more appropriate as compared to the previous version of the ontology.

3.5.5 *The new property containsFocusWord explained*

During the meeting at Bracknell on Feb 8th, 2008, USFD expressed their vision for the Knowledge Fusion and Campaign Discovery component of the system architecture. They proposed that the word frequencies were among the useful elements, identifiable via the Image and Video OCR and Video ASR processes. Word frequencies should also be modelled in the ontology. For this reason we introduced property *containsFocusWord* for the class *Spot* with range string. It is then possible that for each advertisement the most frequent words are listed as values of this property. Similar properties were introduced for the class *Creative* and for *Campaign*. As an example, the advertisement of BT's IT services could be used. The advertisement appeared on Nov. 7th, 2005 in a British newspaper. The words *securing* appeared with frequency 3, *data* with frequency 4, *security* with frequency 2 etc. These words brought the message of the advertisement and bear the attention of the reader, hence the importance of such words in the analysis of advertisements.

To sum up, focus words are words with slight morphological modifications or simply repetitions of the words in the slogan. They serve to reword the main message of the advertising and to amplify its influence on the target group. While logo serves to identify uniquely either the company or its products and brands, the text of an advertisement is not as unique and may vary across different appearances of the same advertisements. However, the text could be shaped in such a way as to deliver a powerful advertising message. The slogan is a phrase repeated thousand times during the campaign so as to become stable in the memory of the target group. Focus words are just rewording of the slogan and by means of creating variations of it, they amplify the power message of the slogan. Sometimes focus words are not morphological variations but simply synonyms or words in associative relationships with the ones in the slogan and as a result they again serve to amplify the main message of the advertisement.

The modelling of focus words in the ontology is not as easy task as it first may appear. There were few important questions which needed solution prior to the final modelling decision. The requirements were to have a modelling of list of words. The questions were how to model the list, should the words in the list have an order or not, should we assign frequency value to each word in the list. In general there were two main possibilities: to model focus words as a class or as a property. If we were to model them as a property automatically the possibility to state their frequency as a property is excluded because in the current state the underlying logic of the ontology is of first order but to state a property of a property requires higher order logic. From the other hand if focus words were modelled as a class then the instances of the class would be numerous and the reasoning which focus word appears in which advertisements would be unreasonably slow and probably poor in performance. It appeared that it would be better to model focus words as a property sacrificing the possibility to state their frequency. Then it would be a straightforward task to list for each advertisement the focus words as values of that property. The tradeoffs would be that this list is not hard coded in the ontology and therefore has no order. But then upon reflection it turned out that the order in the list is unnecessary limitation. The analysis of few real-world advertisements showed that the focus words may appear in the text in any given order. The order of words is significant for slogans only but then it is unnecessary complication to model slogans as a set of separate words rather than as a whole phrase.

The second question was whether or not we are to include stop words as "the", "if" etc. A brief reflection shows that "the" has too high frequency in any text whatsoever and it is not specific for a given advertisement only. Then we either had to stipulate a threshold for the frequency of words in the set, or to give up stop words. In addition if we are not to model ordered set of words, it proves unnecessary to model stop words because they are significant in phrases like "the ultimate driving machine" but not in the set {security, data, desktop, the, reputation}.

The question about lists is an important one. There is a possibility to model lists in an ontology but the current state of reasoners is such that the straightforward and easy solution would be to have a property which specifies a focus word and has a string as a value. In the code, the list of these value is just the repetition of the triple [ad,hasFocusWord,value].

Finally there was a question about the modelling of morphologically different focus words. In the cited example of BT's IT services advertisement, the word security appeared in at least three different variants – security, securing, and secure. It is possible to model this by using something like wild cards.

In the current state of the ontology we stick to the easiest possible solution of the questions discussed above. Focus words are modelled as a property and each separate focus word appears as a value of

that property. The simple repetition of the triple in the code gives unordered list of all focus words for a given advertisement and creative.

To allow stating for each focus words its frequency as well, a simple modelling solution was suggested by the partners of USFD. The solution consists of introducing additional property for the class *Spot*, *Creative*, and *MediaCampaign*. The property takes a string as its value and could be used to calculate the similarity between two different creatives (or spots) by means of performing a cosine similarity metric. For the class *Spot* the new property is named *containsLexicalVector*; for the class *Creative*, the property is named *lexicalVectorOfCreative*; and for the class *MediaCampaign*, the property is called *hasLexicalVector*.

4 MEPCO Ontology – naming and versioning

Second release of the ontology will follow conventions for naming and versioning as described below. The conventions have the purpose to ease the user in deciding which the latest version of the ontology is and hence to ease his/her work with the ontology. As this deliverable is part of the documentation of the ontology any subsequent version should follow the conventions. MEPCO ontology is supplied together with an instance file and the conventions as to the naming and versioning apply in the same way to that file.

4.1 Naming

Naming conventions concern the file name and state the obligatory and optional elements.

4.1.1 Obligatory elements

The ontology file name consists of two elements:

- name of the ontology. It is the first element in the filename. It serves to identify uniquely the ontology.
- extension or file format (the default value is owl). It is the last element in the filename.

Ontology file contains just the name of the ontology followed by the file extension.

Mepco.owl

The instance data file name consists of three elements

- name of the ontology. It serves to specify with what ontology the file is associated and it is the first element in the filename.
- *inst* which hints at the file use - being instance data which extends the ontology. *Inst* is the second element in the filename.
- extension or file format (the default value is owl) . It is the last element in the filename.

Mepco.inst.owl

4.1.2 Optional elements

Optionally one may specify the version number before the extension element. However since the file itself contains such information, this is not mandatory. The optional element may appear only in the instance data filename.

Mepco.inst.2.owl

4.2 Versioning

The code of the ontology, respectively the instance data, must contain enough information so that the user will know exactly which version he/she is using. OWL which is used to encode the ontology offers a straightforward means to express version information by specifying the value of the property version info, where this property is a statement about the ontology. However, we introduce slight changes to be able to specify the date of the ontology release. For the version info the following conventions apply:

- Main version of the file is specified by the first element which consists of exactly one digit.
- Minor version of the file is specified by the second element, also consisting of a single digit
- the month in which the last changes are made is the third element and consists of two digits
- the date when the last changes are made is the fourth element and consists of two digits.

The elements are separated with dots. The valid format of the versioning has the following format:

0.1.10.17,

and reads draft version, first minor version, release data Oct. 17th.

Note that the year is not specified because it is assumed that there is at least one major version per year.

There are other approaches to specify the version info, but the one explained above has been successfully tested in development of KIM platform.

5 Conclusions

This deliverable describes the latest changes in the second release of MEPCO ontology. The ontology was extended so that it could map to output metadata of the different analysis modules in the first prototype. The map would ease the automatic discovery of related creatives and hence the tracking of ongoing advertising campaigns running on media. The changes in the ontology render it richer though not complete as the modeling of the domain is still far from complete. Even though incomplete, the ontology is rich enough to enable description of advertisement appearing in press, television and on Internet, as well as to discover related creatives and thus campaigns. The most important change in the ontology is the addition of properties which model frequency words. Frequency words are exactly the words which bear the main appeal to the target group, grab the attention and create awareness of the advertised product, brand and company. These words, along with the slogan, are markers for the relatedness of the creatives and serve to discover campaigns.

The last chapter of this deliverable explains the naming and versioning conventions which will be followed in any subsequent version of the ontology. In this way the documentation of the ontology is enhanced and a standard in naming and versioning is created.

6 References

- [SA] MC-D61-JRS-09-SystemArchitecture
- [MD] MC-D32-ONTO-04-MEPCO-Design
- [UR] MC-D21-NMRUK-04-UserRequirements
- [MMS] MC-T62-JRS-09-MMS_UserManual
- [Now and McGuiness] Noy, Natasha and Deborah McGuiness, Ontology Development 101: A Guide to Creating Your First Ontology
- [MV] MC-D33-ONTO-03-MEPCO-version-1
- [UM] MC-D642-JRS-01-User Manual

7 Glossary

Terms used within MediaCampaign project sorted alphabetically.

- Campaign** A campaign represents a number of creatives semantically belonging together. A campaign has 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