

# Towards Common Multimedia Ontology Framework – Introducing Modality Perspective

Yulia Bachvarova

Human Media Interaction, Department of Computer Science  
University of Twente, P.O. Box 217, 7500 AE, Enschede – The Netherlands

## 1. Introduction

In this document we briefly present the Modality Ontology [2] which design has been motivated by two main goals: (i) to support the process of selecting the most optimal combination of modalities that can best convey the intended message (ii) to be able to integrate specific knowledge about the strengths and weaknesses of each modality to represent different types of information with knowledge about the structural dependencies or low level features that particular modality exhibits. The latter have already been modeled by some existing ontologies.

At this stage we have developed a lightweight ontology that includes taxonomic hierarchy of all existing modalities. We aim at developing the existing ontology into a heavyweight, axiomatized ontology that can be reasoned upon.

The scope of the ontology we built is mainly determined by the process of selecting the most optimal modality combinations for the particular information that has to be represented.

This process consists of two main sub-processes: selecting those modalities that can best represent the information which has to be delivered (modality allocation) and deciding on the most optimal combination within the set of already selected modalities.

The process of modality allocation requires identifying and representing knowledge about the capabilities of each modality to represent specific information features or types of information. The process of modality combination requires knowledge of the aspects of modalities related to the way they are cognitively perceived and processed, of the features that determine complementarity, etc. In addition, selecting the most appropriate modality combinations requires taking into consideration a variety of different factors, such as the characteristics of the user and the environment, the specific task the multimodal message or multimedia presentation supports, the discourse structure, etc. This makes the selection of modalities extremely knowledge intensive process and thus dependent on the proper modeling of that knowledge.

In the remainder of this document we shortly describe the main levels of the Modality Ontology and building on our experience in its design we postulate our requirements for building Common Multimedia Ontology Framework.

## 2. Modality Ontology

We started building the ontology by firstly deciding on and defining a central operational term. The core concepts which the existing metadata formalisms and ontologies in multimedia currently employ cannot sufficiently and adequately capture the granularity which allows for the scope that we intent to cover with the Modality Ontology. For that reason we have chosen the concept of *modality* as the central operational term and adopted Bernsen's [3] definition of representational modality (as distinct from the sensory modalities in psychology), namely that modality means "mode or way of exchanging information between humans or between humans and machines in some medium". Furthermore the hierarchical taxonomy of modalities we incorporate in the proposed Modality Ontology is based on Bernsen's modality taxonomy which can capture the existing modalities in all the possible levels of granularity [3].

In that respect we consider an important first design decision for building the Common Multimedia Ontology Framework to precisely define an agreed upon meaning of the central operational term (be it modality, media item, multimedia object, multimedia document, etc, or a newly coined term). Such definition will also determine the main knowledge aspects which the common multimedia ontology framework will further extend and elaborate on with its constructs.

We distinguish between two main aspects of each modality – its *content* and *form*. For example the *content* of a picture is described in terms of the concepts depicted in that picture (people, objects) while the form is described in terms of features that characterize the picture as such, for example the property of pictures to be perceived through the visual channel or the fact that they contain highly specific information about the depicted object. Other types of information which also belong to that category are, for example, provenance details. The aspects of *content* and *form* are fundamentally different and, in our view, should be kept separately in the design of Common Multimedia Ontology Framework.

In the Modality Ontology the identity of a modality is described by the PROFILE and the CONTENT sub-ontologies modeling respectively the *form* and the *content*.

Our specific contribution is in the representation of the PROFILE sub-ontology. It has three levels which model three different aspects of knowledge about the form of each modality, namely (i) knowledge about modalities that is related to representing the suitability of different modalities to represent certain types of information, (ii) the cognitive aspect of perceiving and processing each particular modality and combinations of modalities and (iii) the structural dependencies that exist between different modalities and that are directly related to the way meaning is formed.

The *Information Presentation* level models (i). It is built based on Bernsen's Modality taxonomy. We have chosen this taxonomy because it is generative - all modalities are generated from basic properties that uniquely identify each modality in terms of its ability to represent information - and complete - it captures all existing modalities and provides the necessary level of granularity.

The *Information Presentation* level includes the classes *Linguistic*, *Analogue* and *Arbitrary*. The entities that belong to the *Linguistic* category, such as speech and text, have two most notable characteristics – they can abstract and focus. That is, their referential linguistic capacity is detached from the here and now as it can refer to things that are abstract and across time and space. Hence, the linguistic representations focus at some level of abstraction on the subject matter to be communicated without the need to provide its specifics. This stands in contrast to the entities which belong to the *Analogue* category and which depend on how the

subject matter they represent looks or sounds (for example when video is used). Further note that the modalities belonging to the *Linguistic* and *Analogue* categories can function independently in terms of establishing their own domain specific semantic and referential relations and yet they can equally be complementary to each other.

*Arbitrary* modalities are the ones that do not rely on an already existing system of meaning in the use which is being made by them. *Arbitrary* modalities are thus by definition non-linguistic and non-analogue.

The classes of linguistic, analogue and arbitrary modalities can on their own be divided into subclasses based on further differentiating modality features.

The *Perceptual* sub-ontology describes the way each modality is perceived and processed by the human cognitive perceptual channels. We distinguish between the visual, auditory and haptic channels of perception which determine respectively the *Visual*, *Auditory* and *Haptic* classes in the ontology.

Modalities can be further distinguished based on whether they allow to the user freedom of perceptual inspection or not. The class *Static* includes those modalities that can be decoded by the user in the desired order and as long as desired. In contrast, *Dynamic* modalities are transient and do not allow freedom of perceptual inspection. Examples of static modalities are text, map, picture, etc. while animation and video are instances of the *Dynamic* class.

The *Structural* sub-ontology models modality complementarity by determining which modalities are independent, that is, can do substantial representation work on their own and which modalities need other modalities to serve their representational work. Text modalities (written, spoken), for instance are among the most independent ones while graphs tend to be much less powerful in expressing information unless accompanied by other modalities.

For the purposes of clarity we summarize our considerations in the following table.

<b>Scope and usage (annotation, analysis, retrieval, reasoning, personalized filtering, metamodeling, storage etc.)</b>	Support the processes of selecting the most appropriate modality combinations that can represent particular types of information
<b>Media description - Description of the Information Object Covered by the Ontology</b>	All existing modalities hierarchically represented according to Bernsen's taxonomy.
<b>Content Structure Covered by the Ontology</b>	Compositional dependencies between modalities
<b>Content Description Covered by the Ontology – Linking to specific Domains</b>	Does not model the content but the form in terms of suitability to represent certain types of content.
<b>Other Concepts Covered by the Ontology</b>	
<b>Other (upper) ontologies used - modularity</b>	Can be aligned with existing ontologies that represent content thus providing possibilities to model the content – form relationships.

<b>Language - Reasoning Support</b>	
<b>User preferences, specific constraints taken into account, etc.</b>	User preferences, characteristics of the situation, goals of the presentation, discourse structure of the presentation can be taken into consideration
<b>Known Tools using the ontology</b>	
<b>Known projects using the ontology</b>	
<b>Other info</b>	

### 3. Requirements for Common Multimedia Ontology Framework

Building on our experience with the design of Modality Ontology we define the following requirements for a Common Multimedia Ontology Framework.

- Agree upon and provide precise definition of an operational core concept for the Common Multimedia Ontology around which the framework will be constructed. Currently there are a number of operational terms employed by different representation frameworks, for example digital item, information object, multimedia object, multimedia document, to name but a few. Moreover the unclear definition of fundamental concepts such as media and modality has been and still is creating confusion in the fields of multimedia and multimodal systems.  
The core concept of an unified framework has to be defined in such way that it reflects the aspects of knowledge about multimedia the ontology aims to systematize.
- Able to capture knowledge about the suitability of different modalities to represent particular types of information, knowledge about the way they are perceived and processed by the human cognitive system and knowledge about the compositional dependencies that exist between them and that are important for forming meaning. An example of the last point is a compositional dependency that exists, for example, between a map and an icon situated on that map. The map is a substrate which internal structure – points on it correspond to the points of the region it charts – may indicate location of the object depicted by the icon [1]
- Provide possibilities to describe all existing modalities at different levels of granularity
- Model adequately the distinction between description of the content of its core concept and a description of different aspects of knowledge about that concept.

References:

[1] Arens, Y., Hovy, E. and Vossers, M. (1998). *On the Knowledge Underlying Multimedia Presentations*. In Intelligent Multimedia Interfaces. Mark Maybury, editor. AAAI Press/The MIT Press, pp. 280-306, 1993. Republished with introduction in Maybury, M. T. and Wahlster, W. editors. Readings in Intelligent User Interfaces. Morgan Kaufmann Press. 1998. ISBN: 1-55860-444-8.

[2] Bachvarova, Y. and Elouazizi, N. (2005). *Integrating Knowledge about Modalities to a Multimedia Knowledge Representation Framework*. In Proceedings of the Second International Workshop on the Integration of Knowledge, Semantics and Digital Media Technology. EWIMT 2005. Published by the Institution for Electrical Engineering, IEE, London

[3] Bernsen, N.O. (1997) *Defining a Taxonomy of Output Modalities from an HCI Perspective*. Computer Standards and Interfaces, 18:537-553, 1997.