# Media Authoring with Ontological Reasoning: Use Case for Multimedia Information Extraction

**Insook Choi**
Emerging Media Program, Department of Entertainment Technology
New York City College of Technology of the City University of New York
300 Jay Street, V411, Brooklyn, NY 11201

ichoi@CityTech.CUNY.edu

## Abstract

Media Authoring is a versatile practice that can be undertaken as an emerging context for the application of multimedia information extraction (MMIE). The potential value of MMIE in media authoring is related to the need for classifying, storing, locating and retrieving media resources. In this chapter we present the use of authoring data to refine MMIE applied to interactive media, and the use of MMIE data to enhance ontology-based authoring. A proof of concept focuses on nonspeech audio in a multimedia authoring system. Authoring is implemented as path-planning in ontological space; path members are concept nodes that generate queries and return media resources coupled to real-time displays. Ontological organization enables users' orientation to media resources by concept-based navigation rather than by resource type. This approach entails two merits: 1. concept formation takes a significant role in authoring processes and media resources orientation; 2. semantic requirements applied to intelligent authoring in multimedia production are compatible to information requirements for context recognition applied to MMIE.

## 1. Introduction

During the 20th century the majority of media content was generated in one of several well-defined systems—print, cinema, radio, and television—for transmission and reproduction exclusively within the same system. Today, media content is often created by mixing and "mashing up" diverse media types and deployed on diverse devices. Display systems such as personal communications devices can differ considerably from the production systems where media resources originate. And their roles can be reversed: personal devices can generate media in consumer formats that are repackaged and distributed by media syndication (Noguchi 2005). This is a relatively recent phenomenon which is extended by the capacity to exchange media through digital networks and pull technologies.

In an open, multi-platform media ecosystem where amateur-generated resources play side-by-side with those of professionals, the ability to classify, store, recognize, and retrieve media resources is of high value. Semi-automation of media applications through the practice of media authoring enhances the re-usability of media resources in the creation of new content. Media applications can be created that combine pre-selected resources with ones automatically searched and pulled from diverse providers. RSS feeds, Twitter and web-based streaming services provide industrial examples of this process, and indicate a promising future for application-specific and interactive content (Wagner 2009; Hossain 2008; O'Riordan 2008; Blekas 2006). MMIE tools could enhance this production process. However, diversity of media sources and end-user situations increases challenges for structuring common and extensible descriptors.

### 1.1 Organization of this chapter

This chapter demonstrates the use of ontology, a form of well-structured semantic data, as an organizing tool for media authoring. We hypothesize the application of ontology as a robust method for linking media authoring and MMIE. Interactive media is introduced as a context for media authoring, and the practice of authoring with media of multiple types is discussed to identify the relevance of semantic data. A working prototype of a semantic query-based authoring system is described, focusing on the design of ontological data and the use of reasoning to retrieve media resources. A GUI design for navigating semantic data enables queries to retrieve media resources of diverse types in real-time. Analysis of the demonstration system is provided to introduce the relevant application of MMIE and the challenges of robust implementation. Prototype use cases are introduced combining authoring data with MMIE for nonspeech audio.

## 2. Interactive media and media authoring: implications for MMIE

Digital media technology both for consumers and media professionals provides increasing levels of support for interactive and procedural production. Characteristics of interactive media include multiple channels and multiple devices that introduce feedback and control from recipients into the production process. For example: (1) Peer-to-peer transmissions of media resources may be captured as program content; (2) Pre-produced media may be designed as a template or frame for unspecified media resources added during interactive presentation; (3) Media program assembly (editing; mixing) may be located at the point of delivery; (4) Transmission and display systems may be responsive to data about program content.

Interactive media formats enable the configuration of moment-by-moment content to be assembled with respect to users' actions: custom modifications to content are executed before it is transmitted; and content configurations are determined at the point of delivery, on a user's device. Broadcast-era formats are still dominant: pre-produced programs are streamed or downloaded to a user's device. However ISO standards such as MPEG-4 (Koenen 2002) support the transmissions of programs that are not completely pre-packaged, enabling users' preferences, interests and actions to impact the final content.

### 2.1 Authoring

*Authoring* in the context of interactive media refers to the design of instruction sets integrating contents of multiple types, media devices and users' actions. The instruction sets are multiplexed within media content or otherwise transmitted to recipients' devices. An authoring process configures initial conditions, combinations of media resources and conditional procedures for generating program contents incorporating recipients' actions. Authoring instructions not only display or modify content; they assemble content from multiple sources and generate content using purely procedural methods such as graphics and sound synthesis (Delerue 2006).

### 2.2 Media program classification and MMIE

Interactivity creates a new media environment for MMIE, challenging traditional analysis workflow and creating opportunities for new applications such as customized program production. Industrial MMIE applications have been established in response to assembly-line media manufacturing, addressing needs for offline archiving and post-processing for information extraction. Broadcast media provide relatively stable context for MMIE by adhering to well-defined classes of contents that are targeted to well-developed markets. Content classes characterize invariants and meaningful variables in program content: weather forecasts, news broadcasts, sporting events, surveillance, talk shows—each represent well-established patterns of media content presentation, including camera work, lighting, editing, segment duration, figure-ground pattern articulation, and separation of dialogue from music and sound effects. In some respects these templates represent “vocabularies” that broadcast producers have developed to establish observers’ expectations and then repeatedly satisfy those expectations.

Program templates provide essential baseline data for MMIE techniques. These techniques are reliable in relation to the uniformity within the class of media content. This uniformity is related to the broadcasting industry’s model of establishing and maintaining audience expectations, akin to the marketing of commodities based on needs or interests many people have in common. As commodities, broadcast media programs have content variations but their structural templates are relatively invariant. As markets diversify these templates become less reliable at a large scale. With increased capacity for customized contents the nature of media program templates becomes radically altered. Future applications of MMIE depend on development of program classes for interactive media.

### 2.3 MMIE for dynamic and interactive content

Authoring provides production context data relevant for MMIE. The usefulness of the data is determined by the implementation of procedural media processing, modes of user interaction, and quality of media resources. By definition interactive media is dynamic, superseding the notion of a fixed version of program content. Interactive content is dynamically composed in a variety of sequences and layouts and imported from changing sources. Given a blog or YouTube video as a source for interactive content, the generated program content will reflect changes when the web source content changes (Han 2008). Interactive media programs are rarely recorded and archived as definitive documents, only as versions or instances. The workflow for applying MMIE to an archived body of programs is less applicable in this environment due to the high degree of uncertainty regarding the anticipated content and its quality.

We look to authoring requirements to provide a new source of reliable templates for media producers creating interactive programs. In terms of computation and procedural media processing, templates are needed to generate coherent program content that meets users’ expectations while incorporating users’ actions and selections among content options. These templates will be realized in authoring data: computational instruction sets and scripts that encode the authoring decisions.

### 2.4 MMIE Use Case overview

We anticipate that authoring data and MMIE data will be mutually supportive and can be shared. Section 5 presents use cases for these mutually supportive roles. To summarize the baseline use case: Media resources (sounds in this case) are indexed with a structured vocabulary—ontology concepts—then some of the concepts are used to author an interactive media sequence. The generated audio sequence is analyzed using an MMIE tool, with the objective to identify the individual source sounds. Mixtures of sounds in the generated sequence partially confound the audio MMIE tool, such that search results include both correct and misidentified candidate source sounds. The search is refined by comparing authoring metadata with ontological data of the candidate source sounds; false positives are identified when candidate sounds' associated concepts do not match authoring concepts. Thus a concept vocabulary identified with authoring may be used to constrain search space.

We hypothesize that MMIE techniques can be developed to utilize data extracted from patterns in authoring templates as well as patterns in the resulting interactive media contents. Figure 1 shows the proposed scenario: context-rich authoring data provides context data for MMIE. Extracted information in turn can be incorporated as properties of ontological resources and referenced by authoring. This approach suggests a "tuning" cycle for alignment of semantic representations and pattern-based representations of features.

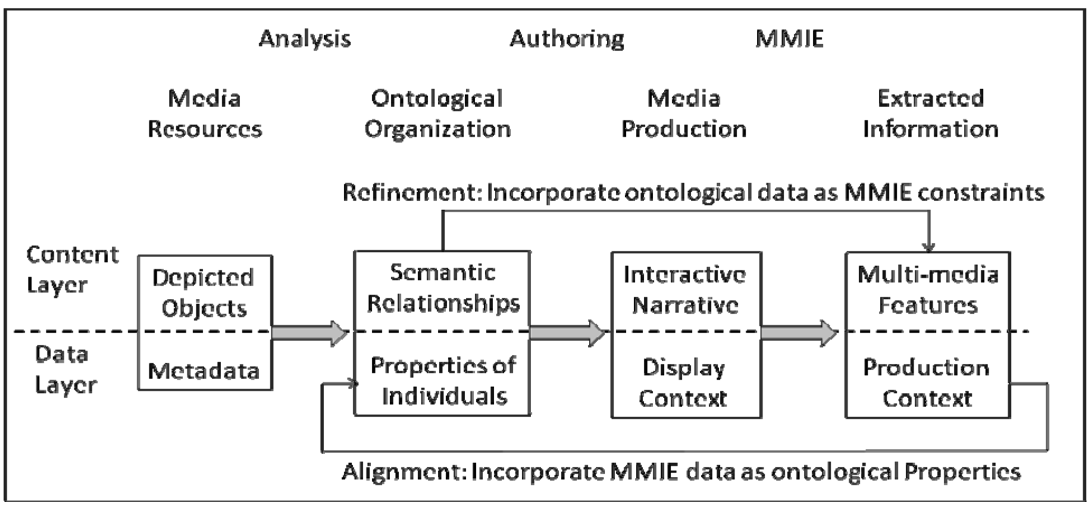


Figure 1: A data layer enables production context refinement for MMIE, and can incorporate MMIE data to align feature extraction with content production.

## 3 Prototype system for media authoring with ontological reasoning

To study enhanced techniques for media authoring, a prototype system was developed to join semantic computation with user interaction. The prototype system tests a capacity for (1) *authoring interactive media*, with (2) an *interactive authoring process*. The first refers to production of conditional procedures for processing media resources while responding to users' actions. The second refers to an authoring environment supported by real-time media processing to display results of procedural conditions. The objective is to support both capacities in close proximity with proper system architecture and performance. Figure 2 represents the architecture of the prototype. Currently-supported media resources include two-dimensional graphic documents such as photographs, diagrams and architectural plans, pre-recorded and procedurally-generated sounds, videos, 3D graphical models, and camera movements in a 3D graphical scene. Semantic representations can be pivotal for linking MMIE across media of multiple types, as discussed in Chapter 9 (Tzoukermann 2009).

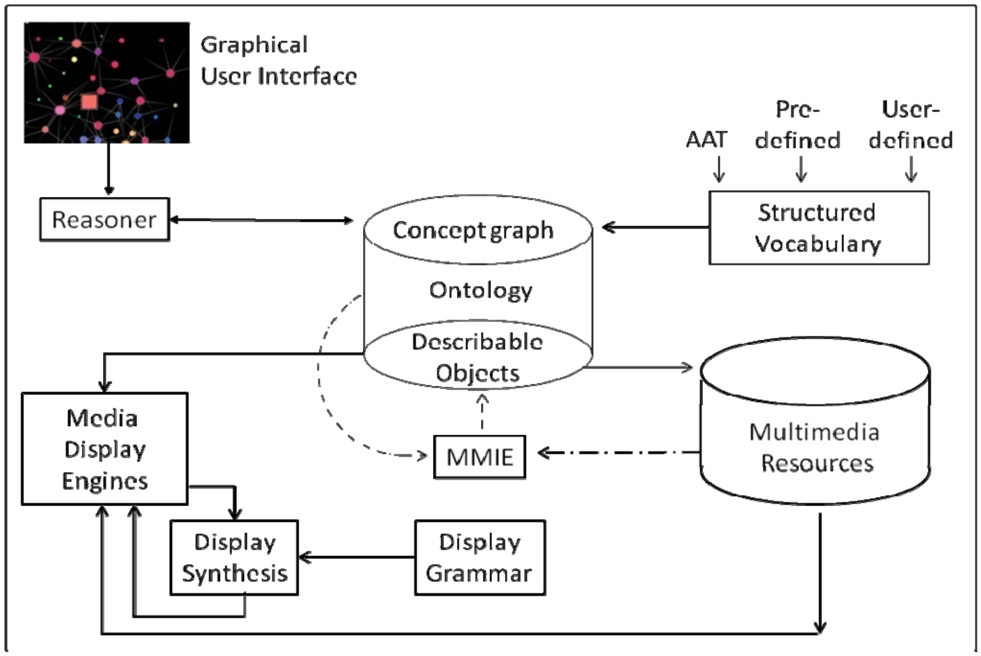


Figure 2: Architecture of the prototype authoring system with MMIE enhancement.

For semantic organization and reasoning we have adopted an ontological data design, discussed in Section 4. The authoring process is formalized as path planning through a semantic (ontological) data structure describing media resources of diverse types. Ontological structure defines a set of logical expressions interpreted as a directed graph of semantic nodes; edges represent concepts' relatedness. Structure is encoded in OWL file format using an open source editing tool (*W3C* 2004; *Protégé* 2009). Paths are sequences of nodes that traverse the ontological data graph. A node can be queried and reasoning identifies semantically related media resources (Tskarov 2002).

**3.1 Graphical user interface for navigating relationships of multiple media types**

A graphical user interface (GUI) supports path-making, and a media scheduling engine displays the results of each query as a real-time mix of media resources. The GUI visualizes the ontological data graph as a 2D network of nodes and edges. The GUI snapshot in Figure 3 displays a limited region of a much larger ontology. Several levels of interaction are defined. Mouse-over a node displays its concept name. Double-click on a node generates a resource query and modifies the display to reveal all nodes that are nearest neighbors. Nodes remain hidden until a nearest neighbor is selected. Nodes are displayed with animated ball-and-spring dynamics, aiding visual identification of relationships. The square node is anchored and can only be moved by direct mouse dragging; it represents a concept's membership in a *query path*, authored as a starting point for concept exploration. The "current location of the user" is defined as the most recently selected node generating the most recent query. Various criteria are applied to hide nodes that become remote from the current query location.

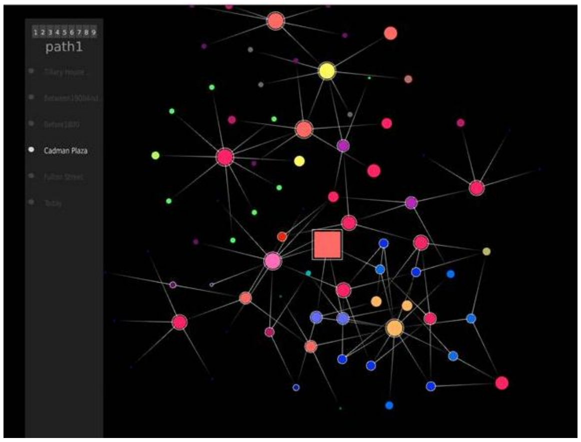


Figure 3: Graphical User Interface of concepts as interactive nodes. Color indicates relationship. Size indicates number of links to a node.

### 3.2 Path-planning and Interactive Authoring

The idea of making paths through a digital document space can be traced to multiple sources, including the Memex proposed by Vannevar Bush in the mid 1940's (Bush 1945). These proposals focus on "trails of documents" using text processing for cross referencing and indexing to achieve more efficient storage and retrieval. Media asset management systems are beginning to adopt these approaches; however computational path-planning has not been adopted for interactive media production.

Our prototype introduces path-planning techniques for interactive media production. Path-planning is computationally robust for combining interactive media and pre-structured media. A path-planning model can accommodate user exploration and improvisation while maintaining linear structures as priorities. Our prototype differs from "trails of documents" proposals by implementing *paths of queries;* paths through concept space generating queries as acts of creative inquiry, creating real-time sequences of composite displays of resources, functioning both as dynamic media content and as semantic navigation feedback to the user. Queries generate sets of related resources; queries organized in paths may be designed to generate linear structures by making a series of selections from media resources that can function in multiple semantic contexts.

In Figure 3 the leftmost vertical array is a path of concept nodes. Path members can be traversed in any order. Each path member displays a square anchor node and a neighborhood expanded for exploration. Selecting a path member generates a query that returns a set of media resources and visualizes the concept neighborhood; exploration of neighboring nodes produces further queries returning related resources.

### 3.3 Interactive Media Use Case Example

A system is configured to respond to queries by scheduling the display of 2D images, sounds and virtual camera movements in a 3D environment. Sounds and images have been entered as media resources in the ontological data set; also entered are virtual camerovements determined by positional data of 3D models. When a query returns one or more 2D images, sounds or 3D objects, separate media display engines receive the addresses of these resources and schedule their display using a sound synthesis system and two image projections, one for compositions of 2D images and the other for 3D scenes.

To create a large-scale working prototype we gathered or generated many media resources related to present-day and historical Brookyln (Choi 2008). The 2D and 3D images are displayed side-by-side in a tiled large-screen format. The GUI is part of a small kiosk. (1) In an example query path the first node is the concept “FultonStreet2000toPresent”. Selecting this concept returns photographs of storefronts; sounds of bus traffic, pedestrians and street vendors recorded on Fulton Street from 2006 to 2008; and a 3D camera movement slowly “flying” (tracking) along virtual Fulton Street with contemporary photographs of real-world Fulton building exteriors applied as texture maps to the building models. (2) Selecting a second path node while these resources are displayed, “BoroHall2000toPresent” introduces new photos and sounds, with smooth visual and audio cross-fades effecting the transition. The 3D camera movement interpolates from Fulton Street to a new position hovering above the model of Borough Hall. Some images and sounds are returned by both queries; these persist in the display across the transition. (3) The third path node “FultonStreet1880to1920” returns the 3D camera to resume a flyover of Fulton Street; however contemporary building texture maps are replaced by historically accurate storefront textures, also the 3D scene now includes an elevated train that ran above Fulton Street in the early $20^{th}$ century. Photographs of hip-hop shops and cell phone vendors are replaced by historical drawings, lithographs, and photographs—including images of the elevated train that provided references for modeling a 3D counterpart. Sounds recorded on Fulton Street are replaced by sounds from a SFX library: horses, carriages, a steam engine, and pedestrians on a boardwalk, synchronized by concept with the images and 3D scene.

Transitions in each media display are computed to dynamically compose image sequences, sound mixes, and combinations of virtual camera movements in the 3D scene. When a user generates a query at the GUI, transitions are effected immediately to provide feedback to the user. Scheduling constraints impose minimum duration between queries. We introduce *Display Grammar* for the configuration of display signal processing and scheduling of multiple resources. Chapter 20 discusses Display Grammar in detail (Bargar 2009).

## 4. Ontological data design for navigating media resources of multiple types

Authoring applied to media of multiple types requires structured access to diverse media resources; it is desirable to develop uniform and extensible authoring procedures rather than tailoring separate authoring for each media type. Ontological data can be designed to support uniform criteria for organizing media resources of multiple types. This task is akin to designing an MMIE meta-program to organize MMIE across multiple types of media. Ontologies are relationships of *concepts* that describe *individuals*; in this case individuals are media resources. “Describe” refers to set membership: concepts describe sets and resources are unordered members of one or more sets. Concept relatedness may be taxonomic subclass-superclass hierarchy or non-taxonomic and non-hierarchical. This flexibility is desirable for authoring both level-of-detail and syntactic relationships.

### 4.1 Dual root structure

Figure 4 summarizes the main components in the media ontology encoded in OWL: a dual root node structure of Concepts and Describable Objects. *Concepts* describe media resources of multiple types, *Describable Objects* include individual media resources and entities depicted by those resources—“Content Objects” in Figure 4. *Content Objects* distinguish specific nameable entities that can be depicted in more than one resource. Content Object types include unique objects detected across multiple resources discussed in Chapter 1 (Das 2009), and named entities discussed in Chapter 11 (MacRostie 2009). *Properties* denote metadata about media resources and content objects, including related objects and quantitative data.

Concepts and individuals may be associated by assertion and by inference. Assertion is a direct assignment of set membership to an individual resource. Inference is a computational evaluation that discovers relationships that may be unknown to the user. Ontological reasoning combines both types of discovery when a query is performed at a concept node.

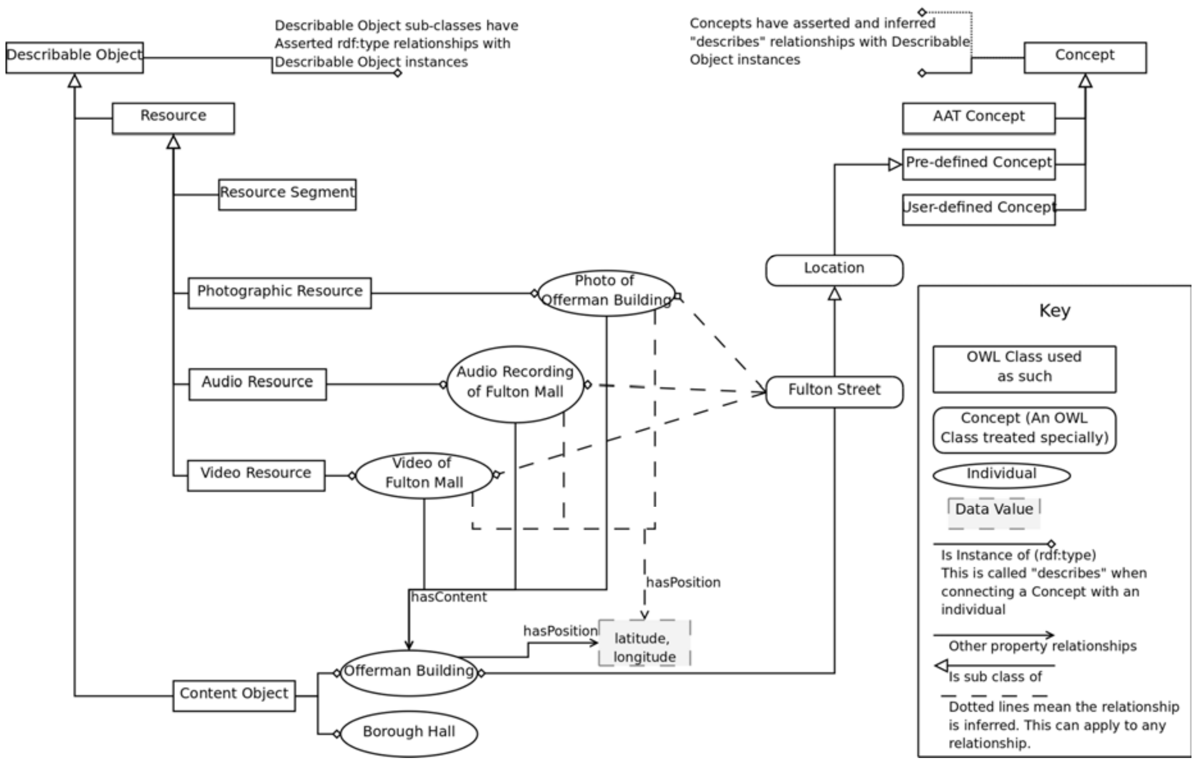


Figure 4: Dual root node structure. Concept and Describable Object are parallel root nodes.

### 4.2 Concepts

Concepts are semantic classes used in logical expressions to show relationships among objects, including other concepts and describable objects. Logical expressions use set membership to represent relationships, which may also be represented as graphs with concept nodes and predicate edges. An ontological data structure may exhibit one of several topologies, including trees (taxonomies), directed acyclic graphs, and cyclic graphs. In media authoring most relationships are syntactic rather than taxonomic.

Concepts may have relationships asserted to specific media resources, however many concepts have relationships only to other concepts. Edges between nodes in the concept graph may be asserted or eliminated without modifications to associated media resources. The dual root design enables concept relationships to be modified without affecting the representations of individual media resources. By minimizing direct dependencies between concepts and resources, changes can be applied to the concept graph without requiring modifications to resources already stored in the system, and resources may be added or removed without modifications to the concept graph.

### 4.3 Complementarity of ontological data and metadata

Authoring can incorporate both semantic data and quantitative data of media resources. Ontological query can reason over both semantic and quantitative relationships, and across media resources of unlike types. Metadata queries cannot; they are limited to individual media types. For example a metadata name-value pair from a photograph might indicate "focal length = 27mm" and from a sound recording might indicate "reverb decay time = 150ms." Taken together these values represent complementary audiovisual data: a visual wide-field perspective a highly reverberant acoustic ambience. This complement is relevant for representations of building interiors. However neither the audio nor the visual metadata provides a structure to define relationships of lens settings to audio signal processing, even though both sound and image metadata may indicate building interiors as a subject. Image metadata can provide focal length information and audio metadata can provide reverb information, but searching on image metadata cannot return an associated sound, nor can sound metadata be queried to return an image.

To address this we apply ontological data design, which can support evaluations of data of unlike types. In the above audiovisual example, the concept "Large Room" is common to both image and sound classifications, and a query can be structured over specific ranges of quantitative values in fields of both sound and image metadata.

#### 4.3.1 Ontological reasoning over resources metadata for MMIE

Ontological data can encompass both a controlled vocabulary of concepts and metadata of individual media resources. Concepts are limited to semantic values and cannot express quantitative values. Metadata can be quantitative but is limited to individual resource types. Concept queries traverse the ontology to retrieve metadata values from individual resources. This indirect access to quantitative values enables comparisons of dates to determine membership in a historical concept, or GPS coordinates to determine geographical location. MMIE data may be stored for individual resources, for example spectral coefficients or pattern recognition data from specific MMIE analyses. Including MMIE data properties in the ontology would enable concept reasoning over known features or patterns identified in media resources.

### 4.4 Properties and Metadata

*Properties* designate data or relationships between individual resources or content objects, illustrated in Figure 4. Media resource metadata is stored as Properties of individual resources, including tags, terms and numerical data. For example the property *hasPhotoSpecific* includes focal length metadata and *has3DModelSpecific* includes polygon count metadata. Properties are either Object Properties or Data Properties. *Object Properties* are extensible and inferable: they are terms and may refer to other properties. *Data Properties* are terminal, not extensible, and may be asserted only, not inferred.

Concept queries may reason over terminal values of Properties including evaluating quantitative data. Properties may be used to store metadata relevant for MMIE such as date, time, or conditions of capture of individual resources. Whereas ontologies describe resources in unordered sets, resources may be ordered using metadata values stored as Properties. By accessing Data Properties a semantic query can evaluate quantitative expressions using values stored in metadata, returning ordered information that cannot be represented directly by ontology.

### 4.5 Design of structured vocabulary

Large-scale organization of a concept structure is difficult to maintain for broad applications, and tends to become taxonomic. To provide an initial semantic order for contemporary and historical resources depicting downtown Brooklyn, a controlled vocabulary was needed that would enable accurate classification of the built environment past and present, interior and exterior, land and water, urban and rural, industrial and agrarian. The Art and Architectural Thesaurus (AAT) (Getty 2005) was selected as a controlled vocabulary that provides both hierarchical and associative relationships among concepts. The AAT does not provide all concepts needed for structuring queries to support media authoring; dates independent of cultural and historical periods, for example, are not included. For the present prototype the AAT provides a semantic anchor for technical accuracy across many types of describable objects, a baseline for building additional concepts needed for media authoring.

*Predefined Concepts* are created to meet the narrative needs of projects. Considerable design is required to minimize redundancy or contradictions with existing vocabulary, to create a limited re-usable non-AAT authoring vocabulary. Combinations of pre-existing concepts may be grouped as *User Concepts*, created on-the-fly by selecting already available concepts and applying operations such as unions, intersections, and filters on metadata values. The intent is to enable queries to be captured, stored and customized to generate results that meet unique needs. Examples of metadata filters in user-defined concepts include dates, GPS locations, polygon counts, and focal length settings.

## 5. Example Use Cases: sound authoring data with audio MMIE

A set of use cases was developed to hand-test the relationship of authoring and MMIE. The tests were confined to sound resources, for simplicity of procedure, and because a small number of sounds when mixed together can generate a large range of complexity for pattern recognition. The MMIE was performed with SoundFisher (Blum 1997), an application for nonspeech audio, particularly effective with nonmusical natural sounds. SoundFisher receives a target soundfile and searches a collection of sounds, returning a list of soundfiles rated in order of similarity. The feature analysis includes distribution patterns and rate of change for loudness, pitch, and spectral brightness. The number of sounds returned varies; a threshold omits sounds with little or no similarity. The SoundFisher algorithm is ingenious and viable over a large range of sound qualities. Only audio signals are analyzed, semantic tags are disregarded. This is critical to SoundFisher’s power and flexibility. The working concept of auditory similarity is unbounded by arbitrary and often comically misleading terminology that populates virtually all professional sound libraries. The similarity function is unconstrained by classifications of sound sources, which typically dominate a listener’s orientation to a sound. SoundFisher is able to identify similarities in sounds of widely diverse origins—sounds that are acoustically similar but generated from dissimilar sources. A listener’s cognitive semantic space expands greatly in the presence of this technological capacity. The SoundFisher classification methodology is complementary to ontological approaches, and provides an excellent enrichment of semantic structure with signal processing analyses. Nonspeech audio is increasingly useful for contextual MMIE, as discussed in Chapters 12 and 16 (Hu 2009; Guerini 2009).

**5.1 Baseline MMIE use case**

The baseline use case applies MMIE to sound resources that are already members of an ontological dataset. Given a stream of sounds generated by an interactive ontology, MMIE is used to identify the sound sources that occur in the stream, and authoring data is applied to refine the MMIE search results, as follows:

1. A query path is authored then traversed interactively, synthesizing a stream of mixed sounds. The sound stream is recorded as a sound file, referred to as the *generated sound*.

2. Temporal data from the path traversal is stored as metadata synchronized with the generated sound, marking the query time of each path node timed from the first query.

3. The generated sound is segmented into 3-second increments, and each segment is used as a SoundFisher search target. For each segment a series of candidate source sound files is returned with similarity ranking. Due in part to sound mixtures in the generated sound SoundFisher may return many candidate sounds, both correct sources and imposters.

4. To refine the search over the candidate sounds for a given target segment of the generated sound, each candidate is queried in the ontology to determine its related concepts. Sounds often belong to multiple concepts. Using the authoring metadata synchronized with the generated sound, the concept set for each candidate sound is compared to the authoring concept active for the current target segment. Authoring concepts that were queried up to 30 seconds prior to the current segment are included, to account for sound file duration overlaps.

5. For each generated sound segment, the candidate sound files are discarded if none of their associated concepts match the corresponding query concept of the segment.

6. For candidate sounds not discarded, common sounds are identified across adjacent segments and compared to the concept authoring metadata to determine overlaps. Durations of the candidate sound files are compared to their contiguous adjacencies across segments. Poor fit of contiguous adjacency to sound file duration demotes a sound.

7. Finally, SoundFisher similarity ranking is applied to resolve conflicts among remaining top candidates, promoting each highest ranked sound at its onset segment.

Figure 5 illustrates Step 3 of this process for one segment of the generated sound, showing the candidate source sounds’ similarity rankings and related concepts.

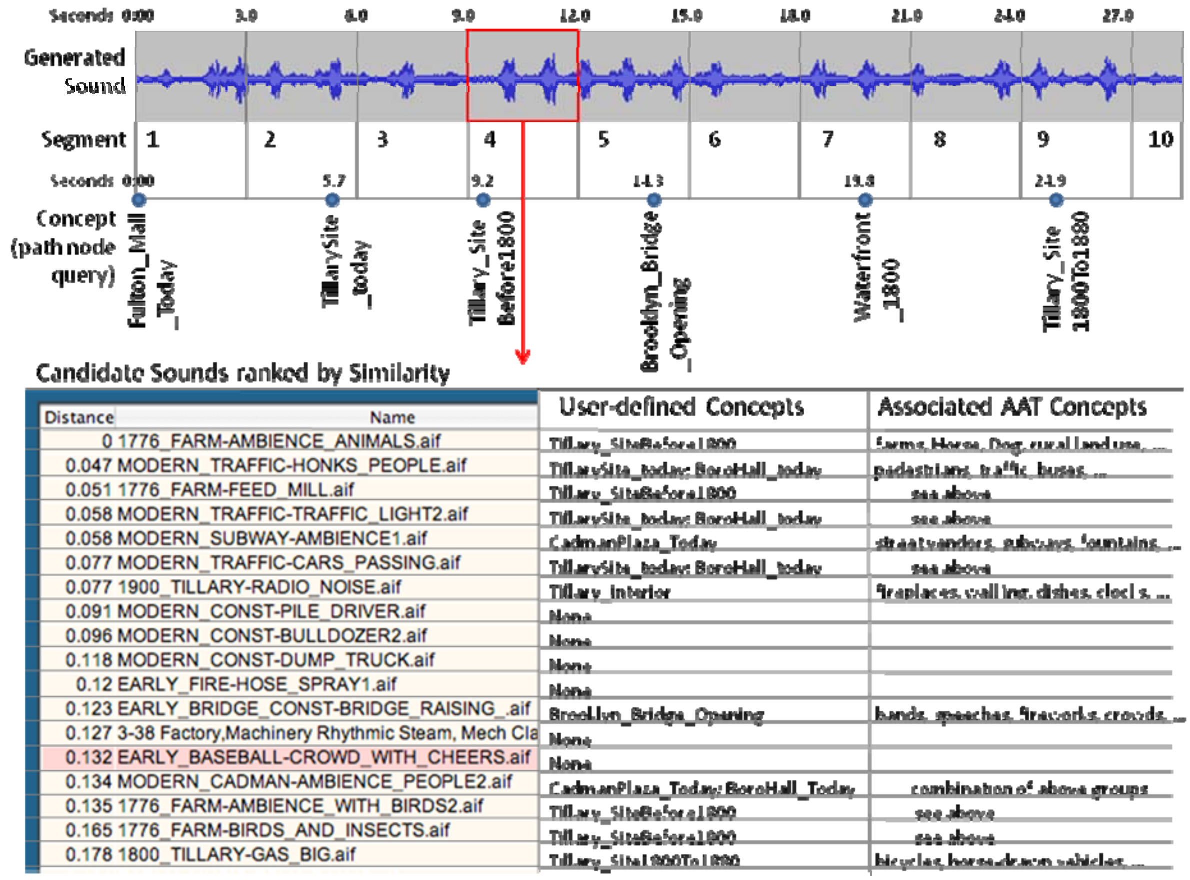


Figure 5: For one segment of a generated sound, MMIE candidate source sounds ranked by similarity. Concepts of each candidate sound are compared to the current path concept.

### 5.1.1 Expanded search

When no candidate sounds are returned for a generated sound segment, a "reverse concept search" technique is applied. (1) From the authoring metadata of the generated sound, the current concept for the segment is identified. (2) Given the current concept, the ontology is queried and all source sounds associated to the concept are returned. (3) One at a time each concept-related source sound is used as a SoundFisher target with the generated sound segment as the sole sound compared to the target. Among the results of these 1:1 tests the strongest similarities are identified as the most likely source sounds of the segment. These results are given a further confidence rating by comparison to results of adjacent segments.

Authoring data can also be used to constrain a search for a target sound within a generated sound. All concepts related to the target sound can be identified from the ontology. The authoring metadata of the generated sound is searched to identify possible locations of these concepts. Then SoundFisher is used to search for target sound similarities only in generated sound regions that correspond to concepts of the target sound.

### 5.2 MMIE data encoded in ontologies

To complete the symmetry illustrated in Figure 1, we enable alignment of MMIE data with ontological data, to complement the refinement of MMIE using the authoring ontology.

SoundFisher similarity results are entered as Data Properties of the individual sound resources returned by each search. We do not have access to the signal processing internal representation of the SoundFisher engine, so for test purposes we adopt a cumbersome approach and enter the rank similarity ordering of each candidate sound with its target sound segment and related concept. The predicate "SoundsLike" was created to support these queries, with additional exposure of the relevant data property. This enables reasoning over the relative similarities of sounds, either with respect to a given concept or disregarding the concept and using only the ordered similarity data. Given that ontological inference cannot return ordered results at the conceptual level, the availability of quantitative and ordered search results is a useful extension for organizing media resources. We are presently experimenting with extended queries such as unions and intersections of sounds according to their similarity. We do not have conclusive results however we hypothesize this may be a way to construct semantic representations of short-duration isolated sound components that can be composed through iterative inferences to sculpt more extended and complex sounds.

### 5.3 Extending ontologies to new media resources

In cases where the original source sound library is not available for MMIE, the authoring data of the generated sound can be used to search surrogate sound libraries. The results of the search will give an indication of similarity to the original source sounds, and can be used to create an extended ontology. If the surrogate sounds are not already members of the ontology, the initial search can be conducted following the baseline use case, applying each segment of the generated sound as a target for a SoundFisher search. Ontological analysis can be performed on the resulting candidate sounds to classify them in the ontology. Results of the SoundFisher analysis can be entered into the data properties of each sound added to the ontology. Ontological inference can be applied to identify additional concept and MMIE data associations, supporting refined and extended MMIE. Given the eventuality of rich extended media resource ontologies, we can expect fruitful results using inference to identify new concept associations to MMIE candidate media resources. In this method the ontological relationships and the media resource data sets are mutually enriched and increased in membership and semantic differentiation.

## 6. Closing statements and future direction

This chapter introduces multimedia authoring as a context for applications of MMIE. Two types of production data benefit MMIE: apparatus variables (device parameters) and production context (location, device and performance conditions). Authoring with ontological reasoning encodes semantic information, includes useful data about production context, and bridges media resources of multiple types. Semantic structure applied to multimedia authoring can provide a formal structure for semantic analysis applied to MMIE.

Our future direction undertakes ontological data design to host a formalized methodology shared by the two practices, multimedia authoring and MMIE. Formalization of shared data protocols will facilitate the use of MMIE as a front-end media production tool. We envision a methodology in three stages:

**6.1 Semantic orientation**

Media authoring can encode semantic information; MMIE optimization requires such information. Media concepts may be used to group MMIE-related data properties to represent well-defined production contexts such as standard media program templates. When MMIE is optimized for a type of media program many assumptions are made about production context. These assumptions may be inaccurate if unsupported by production data. Ontology-based media authoring can host the supporting data, while formalizing semantics of the MMIE process. Semantics describe the structures used to situate meaning, indicating potential situated features sought in MMIE. The identification of a semantic context ascertains a related production context, aiding the selection of MMIE techniques.

**6.2 Semantic detailing**

Content Objects embody semantic details that differentiate the general contents of scenes. Semantic details may be coupled to key features—recognizable patterns in signals that are targets for MMIE. Features may be represented as named, recognizable entities and Content Objects can track them across multiple resources. Content Objects can be grouped to represent combinations of features, stored as MMIE data properties of individual resources. These feature sets may be entities co-located in a single resource or distributed in adjacent resources, such as a sequence of shots in film or an album of images on a web page. Semantic detailing is extensible to *multi-modal features*, such as the co-occurrence of a voice with a face, or a visual scene with a set of environmental sounds (Xiong 2006). Detailing can use concepts to encode cross-media feature sets of MMIE data properties for 2D and 3D visual attributes, 3D model-based attributes, video segments, audio segments, and model-based spatiotemporal attributes such as camera movements in 3D environments. Aspects of multi-modal features are elaborated in Chapters 6, 8, 9, 11, 12 and 16.

**6.3 Feature encoding**

MMIE feature analysis data may be encoded as Properties of Describable Objects. Figure 1 refers to feature encoding as an alignment process; and refinement of concepts by grounding in MMIE data. Concepts could reference classes of MMIE data of known feature types, and could include cross-modal MMIE data extracted from multiple resources. A Content Object-Property relationship can index MMIE analysis data for features identified across multiple resources consistent with their semantic context. Once encoded, the MMIE properties will be accessible to reasoning performed by computational semantics in the authoring process

We anticipate the extension of these techniques beyond the canonical examples of broadcast media, impacting the production and analysis of interactive media. Our goal is to enable the use of MMIE to support shared user-authored media in social networks. New canons and semantically-rich techniques are anticipated in the production contexts created by media producers who are also end-users.

## 7. Acknowledgements

Thanks to Thom Blum and Audible Magic, Inc. for use of the SoundFisher tool. Additional thanks go to Arthur Peters for software development; Robin Bargar for software design and project supervision. Tatiana Malyuta contributed insights on ontological reasoning. Robert Zagaroli, Robin Michals, Mark Skwarek and their students contributed to media resources.

## 8. References

Bargar, R. 2009. Toward Formalization of Display Grammar for Interactive Media Production with Multimedia Information Extraction. This volume.

Blekas, A.; Garofalakis, J.; and Stefanis, V. 2006. Use of RSS feeds for content adaptation in mobile web browsing. In *Proceedings of the 2006 international cross-disciplinary workshop on Web accessibility (W4A).* 79-85.

Blum, T.; Keislar, D.; Wheaton, J.; and Wold, E. 1997. Audio Databases with Content-based Retrieval. In Maybury, M. ed. *Intelligent Multimedia Information Retrieval.* Menlo Park, CA: AAAI/MIT Press.

Bush, V. 1945. As We May Think. *Atlantic Monthly* July 1945: 101-108.

Choi, I. 2008. idBrooklyn: an Interactive Documentary. Live presentation, *Music and the Moving Image Conference*. New York University.

Das, M.; Loui, A.; and Blose, A. 2009. Visual Feature Localization for Detecting Unique Objects in Images. This volume.

Delarue, O. and Warusfel, O. 2006. Mixage mobile. *Proceedings of the 18th International Conferenceof the Association Francophone d'Interaction Homme-Machine.* 75-82.

Getty Trust. 2005. *Art and Architecture Thesaurus (AAT).*

Guerini, M.; Strapparava, C.; and Stock, O. 2009. Audience Reactions for Information Extraction about Persuasive Language in Political Communication. This volume.

Han, Y.; Lee, S.; Kim, J.; and Kim, Y. 2008. A new aggregation policy for RSS services. In *CSSSIA 2008: Proceedings of the 2008 international workshop on Context enabled source and service selection, integration and adaptation; 17$^{th}$ International World Wide Web Conference.* Article No. 2.

Hossain, M.; Atrey, P.; and El Saddik, A. 2008. Gain-based selection of ambient media services in pervasive environments. *Mobile Networks and Applications.* V.13 N.6**:** 599-613.

Hu, Q.; Goodman, F.; Boykin, S.; Fish, R.; Greiff, W.; Jones, S; and Moore, S. 2009. Automatic Detection, Indexing and Retrieval of Multiple Attributes from Cross-lingual Multimedia Data. This volume.

Koenen, Rob, ed. 2002. Overview of the MPEG-4 Standard. *International Organisation for Standardisation (ISO) JTC1/SC29/WG11 Coding of Moving Pictures and Audio*.

MacRostie, E.; Watson, J.; and Natarajan, P. 2009. Extraction of Human Language Content in Video. This volume.

Noguchi, Y. 2005. Camera Phones Lend Immediacy to Images of Disaster. *Washington Post.* Friday, July 8, 2005: p.A16.

O'Riordan, A. and O'Mahoney, M. 2008**.** InterSynd: a web syndication intermediary that makes recommendations. In *Proceedings of the 10th International Conference on Information Integration and Web-based Applications & Services*, 299-304.

*Protégé Ontology Editor Knowledge Acquisition System*, 2009. http://protege.stanford.edu/

Tskarov, D. and Horrocks, I. 2006. FaCT++ Description Logic Reasoner: System Description. *Proc. of the Int. Joint Conf. on Automated Reasoning* (IJCAR 2006), 292-297.

Tzoukermann, E. et al. 2009. Semantic Multimedia Extraction using Audio and Video. This volume.

*W3C: OWL Web Ontology Language Overview*, 2004. http://www.w3.org/TR/owl-features/

Wagner, E., Liu. J., Birnbaum, L. and Forbus, K. 2009. Rich interfaces for reading news on the web. *Proc. of the 13th International Conference on Intelligent User Interfaces*, 27-36.

Xiong, Z.; Regunathan, R.; Divakaran, A.; Rui, Y.; and Huang, T. 2006. *A Unified Framework for Video Summarization, Browsing and Retrieval*. Burlington, Mass.: Elsevier.