Toward Semantic Interoperability in Description, Sharing and Retrieval of Media Objects

Youngchoon Park, *PanKoo Kim, Forouzan Golshani, Sethuraman Panchanathan,
*Collage of Computer Engineering, Chosun University, GwangJu Korea
pkkim@mina,ehosun.ac.kr
Department of Computer Science and Engineering
Arizona State University, Tempe, AZ 85287-5406
{ycpark, golshani, panch}@asu,edu

ABSTRACT- In depth of problems and potential solutions for heterogeneities in content description (annotation) is presented in this paper. As a result, we present an efficient media content description scheme with a clear emphasis on the interoperability among various applications with respect to semantic and syntactic content descriptions. Such concept-based descriptions are fundamental to sharing, indexing and retrieval of media objects. To minimize the heterogeneity among text-based semantic annotations, a large ontology with strong concept naming rules is introduced. This forms the cornerstone of a conceptual-graph based annotation framework, Descriptions are then implemented as part of a unified content description scheme called UCDS, which is based on an extended XML syntax. The semantic contents of media objects are annotated with concepts and their relations. The proposed scheme is evaluated through an implementation of the semantic content retrieval of media objects that are annotated based on the concepts and their relations, Details of indexing and retrieval methods are presented in a systematic way.

1. Introduction

The large-scale proliferation of multimedia data necessitates the use of sophisticated techniques for accessing the information based on the content, Indexing and retrieval of visual content from video databases requires sophisticated audio-visual feature and content extraction techniques, content description methods, multi-dimensional data indexing methods, and efficient similarity measurement techniques. To satisfy the demand, recent research has produced many novel techniques for content-based visual information retrieval (CBVIR). CBVIR has emerged as a promising yet challenging research area in the past few years.

As a consequence of the research activities, several prototype systems and commercial retrieval engines including, QBIC [9], VisualSEEk [19], VideoRoadMap [13] and MP7TV [16] have been developed and used. However, as yet, no existing system is capable of complete understanding the semantics of visual information, even though matching based on generic features such as color, size, texture, shape, object motion are well within the state of the art [9][12].

Regardless of media type, users tend to find media objects based on their perceptual understanding of the media content [16]. Such perceptual understanding can be a semantic recognition of a localized visual object in an image, an emotional perception of a certain sound clip, an abstracted summary of a news video clip or a particular semantic event captured from a movie clip. These facts have been explored in our human factors study in visual information retrieval. To enable conceptual level querying, the semantics of media content should be captured and represented in computable form. In the context this work, semantic refers both the "meaning" and "description purpose" of a media object. However, current content-based media retrieval techniques are not mature enough to answer semantic related queries. This is due to the difficulties in extraction of semantic information from low-level media data.

The most frequently employed semantic annotation mechanism is to use keywords or to use simple natural language sentences. However, our recent human factors study reported in [15] shows that keyword-based visual content annotation and retrieval requires a sophisticated handling of keywords since, two people may use the same term with probability < 0.28 to describe or retrieve information. This term mismatch has created difficulties in information retrieval.

One of the simplest solutions for the problem of term mismatch is to use a set of predefined vocabularies in both description and retrieval of semantic content of a media. However, it is almost impractical to force users to use the same term in indexing and retrieval because, different user-groups often interpret information in different ways based on their own domain knowledge or contextual space. Different user groups produce different conceptual interpretations for the same piece of information[16]. It is also governed by a user's perception of information content in a particular context [18]. These differences indicate that it is not sensible to force users to use identical sets of terms for information retrieval. Depends on application domains, and users, there are number of valid content descriptions for a single media object. Typical users would expect the information system to interact not at the data or feature level, but the information or knowledge level retrieval. In fact, future content-based multimedia access systems should provide a method of retrieving a set of media objects from multiple data sources based on user preferences at the level of media semantics as well as media features. Traditional indexing and retrieval techniques are can be efficiently used when the number of relations and relation types are predefined.

However, such static modeling of relations may not be an efficient method in semantic-based retrieval of media due to the fact that the number of semantic relations is too large and the relations have to be managed dynamically.
A promising solution for providing smart retrievals (semantic/knowledge level retrievals) of media object based on the content is to use concept-based indexing and retrieval [16]. Concept-based indexing and retrieval of media content attempts to explore beyond the standard keyword, and audio-visual feature based indexing approaches, which use simple counting of the words or media features from the given user query. Concept-based indexing technique uses knowledge of conceptual interrelationships among concepts to find correspondences /semantic relations between the concepts in user query and that occur in media contents. Figure 1 depicts a scenario of concept based indexing and retrieval. In this scenario, users may access all types of media content from a unified conceptual view. For instance, a sound clip of an airplane landing, a video clip of an airplane landing may be retrieved at the same time. By consulting conceptual relations kept into knowledge base or ontology, various pieces of information related to the media content can be supplied in semantic retrieval.

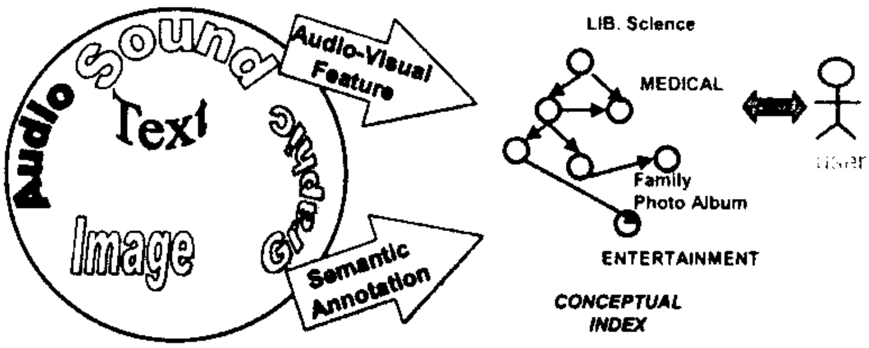

Figure 1: Concept based media indexing and retrieval

Concept-based indexing and retrieval requires a knowledge base that defines semantic type hierarchies and semantic relations among concepts. Advanced knowledge base may contain inference rule for semantic mapping from data to a concept. Various knowledge bases with knowledge description mechanisms [8][6] can be applied for this purpose. One kind of knowledge representation and organization is an ontology. In knowledge/ information management systems, an ontology provides the primitives needed to formulate queries and semantic resource descriptions. In addition to this, ontology may provide semantic interoperability over heterogeneous semantic descriptions by providing a set of standard vocabularies and knowledge about semantic relations among vocabularies [8][10][11].

One of key challenges dealt with in this research is to provide a semantic level interoperability among heterogeneous semantic descriptions. A scheme called UCDS is presented to resolve the problem of the semantic interoperability by providing a unified content description scheme that is expressive enough to convey intended human interpretation of the media content. UCDS has been inspired by various standards and research activities including MPEG-7, RDF (Resource Description Framework), Knowledge Integration Environments, Formal concept analysis, ontology-based information retrieval, and content-based audio-visual information representation. In short, UCDS is a scheme specification for description, sharing, and retrieval of the media contents over heterogeneous media repositories. We will also present some operational techniques for indexing and retrieval of media objects based on the their concepts and a large number of semantic relations.

2. Description, Sharing and Retrieval of Media Content

Our view of a media content description is **a computable representation of conceptualized media content based on a given context.** Throughout the history of information retrieval, context has been either ignored or considered as "something that already exists in some where". Lacking of contextual information in content-based information indexing and retrieval produces the problem of domain subjectivity and user subjectivity. Such problems are due to the limited expressiveness of languages that are used for communication between users or between users and machines and lacking of contextual information to understand an intension of description and retrieval. Therefore, description view and retrieval view should clearly be identified and incorporated in a description as well as in a query.
Let us consider a simple information repository denoted by E, as shown in Figure 2, an annotator A sees an image M, and describes the content of M using a description language. Suppose Ia is a computable expression of media M by an annotator A. Suppose a user, B wants to find the M, then B has to compose a query with a language expression Lb that is can be understood by E, where Lb denotes the B's mindset (or query) on the media M. In general, a formal description La and query expression Lb are not the same with respect to the syntax and semantic, therefore, annotator A is required to have a knowledge of what has been described and to know how to retrieve M. A user, B may not be required to know what was the actual content of M, but he has to know how to formulate a query expression Lb to find M. Suppose user A uses "sky and cloud" as a keyword for M, while user B uses "nimbus", a dark gray cloud baring rain, as a query (suppose user B has a particular type of cloud in his mind). In this situation, user B may not able to find M, even though "cloud" and "nimbus" are conceptually similar to each other. The reason is that annotator A's conceptual understanding of media M is more abstracted than that of user B.

The next class of heterogeneity is heterogeneity in description and query languages. An ideal situation is when user's conceptual interpretation of M denoted by Ia is identical to a translation of Lb. However, almost certainly, they will not be entirely identical since there is no conservation of the *context* of content description during translation from one language to another. In summary, major problems in media content description, sharing and retrieval are as follows:
i. Annotator and retriever are at the different conceptual spaces (heterogeneity in conceptual view)
ii. Limited expressiveness of description languages
iii. Limited expressiveness of query languages
iv. The description language and query language are not identical
v. Lack of contextual information in description and retrieval
vi. Lack of description validation and verification, in particular, at the semantic level.
In the subsequent sections, we will propose our solutions for addressing the problems listed above. Reported solutions are expressed using UCDS.

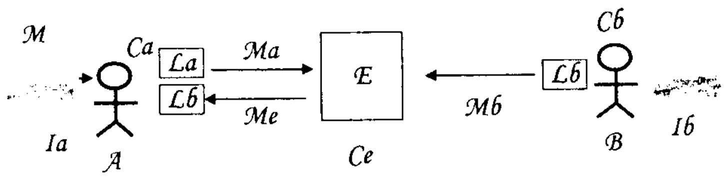

Figure 2: A simple information retrieval environment

2.1. Heterogeneity in description and retrieval of syntactic content

A number of efforts on automated extraction of syntactic content such as color histogram, curvature histogram, etc., from a media object have been reported. Selection and usage of these automated content extraction methods highly depend on the application domain, and the media type. A good summary of such techniques may be found in [11]. In addition, standardized description schemes such as MPEG-7 are essential for information exchange and integration.
Heterogeneity occurs at the various description levels including description term level, schema level and description structure level. In fact, MPEG-7 addresses the minimization of heterogeneity at the term level and schema level by standardizing the set of description elements (e,g., color histogram and shape descriptors) and by providing description type definition, including description syntax and schema, However, allowing users to define their own description schemes with the DDL (description definition language) may introduce new problems stemming from heterogeneity at the schema level.

The syntactic content of a media object is characterized by a set of descriptors in MPEG-7. Depending on the application domain, a different set of descriptors and description schemes may be generated. Such extensibility in MPEG-7 requires a certain level of intelligence in indexing and retrieval of heterogeneous MPEG-7 documents. There are a number of situations where heterogeneity may occur. Given two description schemes, namely Dl for a MPEG-7 document MDl and D2 for MD2, there is a case where Dl and D2 refer to the same syntactic content, however, their description type definitions (DTD) are not identical to each other, This type of heterogeneity can be easily eliminated by using simple translation of type definition. The next case of heterogeneity is that Dl and D2 are two distinct descriptions and they are the only descriptors in MD 1 and MD2. To retrieve MD1 based on the D2, a description corresponding to D2 should be identified from a media object. For the purpose of information integration, such generation of new description is always necessary. Syntactic description translation requires the information about the convertibility between two descriptions. Based on the input data requirement, convertibility can be classified into (1) directly convertible and (2) indirectly convertible. The first case is that there is a method, which can translate a description to the other without requiring the original input data (i.e,, HSV histogram vs. RGB histogram). The second case requires the original input data. In fact, the second case is not a translation but it is actually generation of description. Most of syntactic content descriptors belong to the second case. Therefore, it is necessary to access low-level media data in indexing and retrieval despite the existence of MPEG-7. This physical media referencing may cause computing overhead in indexing and retrieval. In the media indexing phase, at a minimum, we require four procedures, namely, locating the physical media, transmission of a media, decoding the media, and analysis of the media. Therefore, it is beneficial to include an approximated version of the object with an acceptable quality in the syntactic content description. This will lead to a more efficient physical media accessing mechanism.

In our work, a media object representation scheme based on polygonal mesh is used, It incorporates syntactic descriptions for the purpose of indirect description conversion and delivery of media content.

2.2. Heterogeneity in description and retrieval of semantic content

In general, heterogeneity in description and retrieval appears when multiple application domains, many annotators, and multiple users are involved in multimedia information management. For instance, descriptions of plant images in a plant encyclopedia and descriptions of the same images in a National Geography database application may not be the same. In particular, each application may use different descriptions at different levels to describe a plant occurring in an image. The generic name of the plant and color of the plant might be sufficient for a National

Geography image database, however the scientific name of the plant (i,e., the botanical name), color and shape of leaves (e,g,, rose vs. rosa or pendulina) will be essential in plant encyclopedia.

Let us consider another example. A description term, "behind", may refer to a state of "toward the rear" in basketball video database, or, it may refer to "a step with the free foot crossing behind the weighted foot" in a dance video database. In this situation, detection rules (or understanding rules) of "behind" from a basketball video are not identical to the rules for detection of "behind" from a dance video. From the retrieval point of view, there are two distinct heterogeneity types in conceptual views. The first case happens in a single conceptual space. The second case occurs when multiple contextual spaces are involved, as presented in Figure 3. The left side of Figure illustrates the first case. A database application holds all descriptions that are in the semantic range between a and b (annotator's conceptual space), where a is at a more abstracted conceptual level as compared to b. If a user posts a query that is not at the same conceptual level where a target description resides, then a semantic query relaxation is needed to retrieve a set of media description.

Another type of semantic mismatch occurs when a retriever and an annotator are in the same level of conceptual space and refer to the same semantic entities with different description schemes. Suppose, both the annotator and the retriever use the same notion of color histogram in image description, however, the annotator uses a normalized histogram in description and the retriever use ordinary histogram in query formulation. A similar situation frequently occurs in keyword-based information retrieval where a semantically organized set of keywords (e.g,, an electronic thesaurus) or controlled vocabularies with semantic relations have been used.

The last type of semantic mismatch (shown in the right side of Figure 3) occurs when annotators and users are in different conceptual spaces. For example, suppose a user who is in the domain of dance education, wants to find a video clip that contains motion description called "behind" from a domain of sport video. However, the meaning of the description symbol "behind" in dance domain is not the same as the meaning of "behind" in a sport domain. In fact, as stated earlier, "behind" in a dance domain is a step with the free foot crossing behind the weighted foot, Without haying information about conceptual space where a description is valid, we cannot ensure the correctness of classification and retrieval of the semantic contents of multimedia.

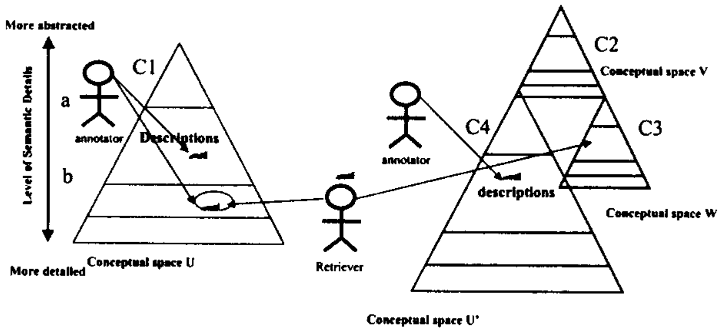

Figure 3: Possible semantic mismatches in description and retrieval

There are two possibilities for resolving this problem, The first is conceptual space integration through partially automated tools. The integration produces a unified conceptual hierarchy, and therefore, allows a unified view for conceptual space management. However, dynamic changes of conceptual space (i.e., frequent introduction, modification or deletion of concepts, relations, and rules) would cause a serious problem. The alternative solution proposed in our uses a conceptual type hierarchy manager, which acts as a resource manager maintains semantic relations among conceptual structures.

3. UCDS for Content Description and Markup

In this section, we will introduce UCDS (Unified Content Description Scheme) which is an XML application for description, sharing, and retrieval of multimedia information with particular emphasis on semantic content description. For a given application or domain, the process of content description for a media object begins with the identification of content elements and their relations, leading to a model of the basic aspects of the content. One of the distinct features of UCDS is the provision for a view-based description mechanism, similar in spirit to database views. This is a relatively new idea, which has not been used in content-based information management, in particular content description schemes.

In UCDS, media contents are described based on media objects and their relations. A media object is the fundamental entity in description of media content at various levels, which can be defined along with different semantic and syntactic dimensions. For example, a region in an image, a moving object in a video or a sound clip is considered as a media object. Even an entire image or video may be a media object in UCDS. In addition, different media descriptions may be used for different levels of media objects.

A description of a media object may include a number of description elements. Each media description element represents an atomic property of a media object such as color property, shape property, semantic property, etc. A description element may have a non-recursive equation of arbitrary rank n. In general, it may be a simple or complex data structure. Relation description among media objects may present various semantic or syntactic relations. As an example, relations of objects may represent spatio-temporal arrangement of objects or semantic relations. Relations in UCDS are represented by n-ray predicates.

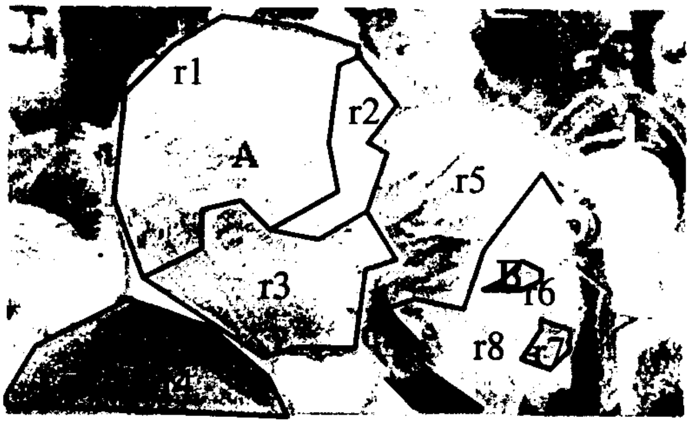

Figure 4: Multi-level annotation of an image captioned: "US President Clinton's spent busy day campaigning for his wife Hillary"

Let us consider the image shown in Figure 4. Depending on the description level, there may be a number of media objects. An entire image can be just one media object. In this case, obviously, no internal relations exist. In a region-based description level, the image contents can be described by a set of regions (based on their associated visual properties) and their spatial relations, Other levels of description are also possible. For instance, a media object called "Former US president Clinton" denoted by "A" in Figure 3, may be represented with a set of regions (in the example, r1, r2, r3 and r4) and their spatial arrangements with shape properties of "A" in semantic object-based description. In this level, there may be various semantic relations among media objects, e.g., Clinton has a wife Hillary Clinton, denoted as B with the property wife(A, B), and B has a husband A denoted by husband(B, A).

3.1 Metadata Description

Metadata plays more an important role in managing multimedia information than does the management of traditional structured data or textual information retrieval [18]. In addition, content-based media search may not be essential for each and every retrieval, since applications vary considerably and the retrieval constraints are generally different in each case. As an example, we do not need content based access for retrieving copyright data or publisher information. There are a number of metadata specifications for media management, including Dublin Core[3] Elements Set, MPEG-7 Metadata DS [14], among others. UCDS does not specify any standard metadata tagging set, however, UCDS allows incorporation of the existing metadata description sets.

Metadata description may include description of content encoding views and other metadata description components. These descriptions define the view of description. For instance, descriptions of a conceptual space illustrate the semantic space (domain of semantic encoding) where semantic descriptions of a given media are valid. View definition is part of metadata descriptions.

Several types of metadata are necessary in media description to capture a fair amount of content information as well as the semantics of the application area. Added to these is a variety of contextual information. An example of metadata description in UCDS, relating to the picture presented in Figure 4, is presented below.

```
<metadataDS>
    <!-- Use dublin core element set -->
    <!-- importing the existing metadata description -->
    <use-metadata-specification xmlns:DublinCore = "http://foo.org/dublin/dunlin.dtd">
          <DublinCore:Description about=" http://foo.org/demo.mpg">
<DublinCore:title> US President Clinton campaigning </DublinCore:title>
                              <DublinCore:creator>Mr. Foo </DublinCore:creator>
                              <DublinCore:description>
                                          US President Clinton spent busy day campaigning
for his wife Hillary
                              </DublinCore:description>
                              <DublinCore:publisher>MSNBC</DublinCore:publisher>
                              <DublinCore:date>09/21/2000</DublinCore:date>
                              <DublinCore:rights>NBC</DublinCore:rights>
          </DublinCore:Description>
    </use-metadata-specification>
    <!-- use some of mpeg-7 metadata specification -->
    <use-metadata-specification xmlns:MPEG-7 = "http://foo.org/mpeg7/mpeg7xm4.dtd">
<MPEG-7:MediaLocator>
          <MPEG-7:MediaURI> http://foo.org/demo.mpg </ MPEG-7:MediaURI>
          <MPEG-7:MediaTime>
                <MPEG-7:RelTime>PT3S</MPEG-7:RelTime>
                <MPEG-7:Duration>PT10S</MPEG-7:Duration>
          </MPEG-7:MediaTime>
</MPEG-7:MediaLocator>
    </use-metadata-specification>
</metadataDS>
```

Metadata descriptions are expressed within two statements <metadataDS> and </metadataDS>. The statement, <use-metadata-specification> is for defining a domain/application specific metadata description set. It adds the vocabularies, structures, and relations of the specified metadata description set to the current description. To incorporate multiple description schemes in a single description, we may use XML namespace recommendation rules. The above example illustrates the inclusion of existing metadata specification standards. When domain-dependent metadata is required, users may create a metadata description set and may indicate that there is a user-defined metadata set with the tag <use-metadata-specification>.

3.2 Media Content Description

Media content descriptions in UCDS describe the actual content of a media. UCDS allows two distinct types of descriptions, namely syntactic content description and semantic content description. Syntactic content refers to the content information that can be directly accessible from low-level media data with a certain degree of automation and without recognition. Semantic content refers to content information that can be directly or indirectly accessible from low-level media data with contextual information and recognition.

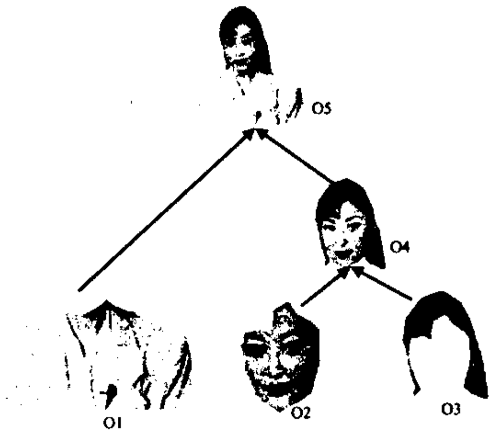

Figure 5: A media object, O4, is constructed from the two media objects, O2 and O3.
A media object, O5, is constructed from O1, and O4= {02, 03}

A media object is the fundamental description entity in UCDS. A media object can be described at various levels along with different semantic and syntactic dimensions. For example, a region in an image, a moving object in a video or a video clip is considered as a media object. As an example, from low-level image analysis, we may have three distinct media objects namely O1, O2 and O3. See Figure 5. If a user gives a symbolic name, "upper body" for O1, "face region" for O2, and "hair" for O3, then he/she may construct a new media object, 04 by using O2 and O3. Let us assume that 04's annotation is a "akyo's face", O2 and O3 may be represented with color attributes and symbols while, O4 is described with shape, name, and spatial relations of O1 and O2. This example shows different levels of media object description scenario as well as a composition hierarchy of a media object. By only looking at the syntactic construction of a media object, the skeleton of media object description for Figure 5 may be expressed as presented below.

```
<MediaContentDS>
        <MediaObjectDS ID="O5">
                       <con>
                                     <MediaObjectDS ID = "O4">
                                     <con>
                                                   <MediaObjectDS ID = "O2" >
                                                                 <SyntacticDS>

        <!—Audio-visual content description of O1 -->
                       </SyntacticDS>

</MediaObjectDS>
                                                   <MediaObjectDS ID = "O3" >

        <!—Audio-visual content description of O2 -->
</MediaObjectDS >
<!—Audio-visual content description of O3 -->
</con>
</MediaObjectDS >
                                     <MediaObjectDS ID = "O1" >
<!—Audio-visual description of O4 -->
</MediaObjectDS >
                       </con>
                       <SyntacticDS>
                       <!—Audio-visual description of O5 -->
                       </SyntacticDS>
</MediaObjectDS >
</MediaContentDS>
```

Descriptions of media contents are expressed between the two statements <MediaContentDS> and </MediaContentDS>. The statement <con> indicates that the current media object being described is constructed from media objects listed within a scope. In fact, the above description illustrates a composition hierarchy of media object O5. Individual descriptions of a media object are expressed within two tags <MediaObjectDS> and </MediaObjectDS>, Attribute ID in <MediaObjectDS> is a unique identifier of a media object in a UCDS document.

3.3 Description of syntactic content

Syntactic content of a media will be described within a scope defined by <SyntacticDS> and </SyntacticDS>. Each description element consists of (1) description type, *T*, (2) a description name, *N*, and (3) actual instance of description, *I*. Formally, a syntactic description element D_{syn} is defined as follows:

$$D_{syn} = <T, N, I>$$

Input data descriptions play an important role in syntactic content interoperability. Generally, translations or generations of a descriptor from one media content descriptor to another requires access to the physical media object. Consider two arbitrary images Il and I2, Suppose II is represented as a composition of regions with associated color features, and I2 is described with a global color feature, If a user wants to search based on the region, then there is no mechanism to compare Il and I2 directly, since they are described with different criteria. To compare them, region properties of 12 will have to be converted for matching.

3.4 Description of Semantic Content

Among UCDS's distinct features is the method employed to encode and describe the semantic contents, which is different from the currently existing description schemes. We use a conceptual graph formalization to describe the semantic content of a media object. In a conceptual graph, a concept is represented with a circle. A labeled arc between two concepts represents a semantic relation. Most labels and concepts typically used in conceptual graph formalisms are developed in an ad-hoc manner with such relationships as part-of, has-histogram, etc. These ad-hoc creations of concept names and relation names cause a problem in semantic matching. For instance, "has-part", "part", and "has-element" can be considered to represent the same semantic relations. To resolve the problem, UCDS only permits to use a word defined in one of the ontologies. The participating ontologies must be identified in the metadata-description section. In this way, computer-aided description tools may establish proper naming environments. In UCDS, nouns and predefined symbols (reserved symbols in UCDS) are allowed as relation names. Concept types and relation types must be defined in the ontologies. With these restrictions, a UCDS processor can validate a semantic description.

In the semantic annotation phase, an annotator uses vocabularies that express his semantic understanding of the media content in his conceptual view. The notion "view" in UCDS is characterized by (1) a set of vocabularies and (2) a set of rules for content interpretation. Therefore, intended meaning of a semantic annotation is characterized by a conceptual space, a concept type and its symbolic representation. An ontology in UCDS is employed to define a domain of semantic interpretations (view) and concept types. To encode conceptual element (which may represent either a concept or a relation) precisely, UCDS requires three distinct types of information, namely:

1. an ontology,
2. a concept type defined in (1), and
3. user-defined referent.

A concept type or a relation type in UCDS is defined as a conceptual inherent path from a unique empty concept (root concept of all concepts) to the target concept.

A conceptual element or individual description of an atomic concept in UCDS is defined by a conceptual ontology, a concept identifier, a concept name (referent), and concept type, where concept type is an element of the given type hierarchy. A relation in UCDS is defined by a relation name (referent), a relation type, and an argument list, where relation type is a vocabulary in the ontology. Based on these restrictions, a concept, say "an anchor", can be expressed as follows:

```
<SemanticDS>
<ConceptDS ID = "423215-1030-32">
      <use-ontology
uri = "http://foo.org/wordnet-ontology/wn-ontology"
use-prefix ="wordnet"
version-info = "1.6">
                              <concept-type>
                    ϕentity@life_form@person@communicator@reporter@television_reporter
                              </concept-type>
          <referent>
                                   anchor
                              </referent>
                    </use-ontology>
<ConceptDS>
</SemanticDS>
```

The statement <use-ontology> adds semantic encoding view of concept type to the current conceptual element description. In fact, it defines the source of ontology that will specify a set of possible concept types and relation types for the current description. In addition, an ontology identifier may be defined by using "use-prefix". It is designed to allocate an identifier for the ontology being used in the current description. The uri in <use-ontology> illustrates the physical location of the ontology. If a UCDS application (e.g., index application) obtains ontology information by using a document object model (DOM), then it will make contact with the physical location and create a channel to communicate the ontology source for future use.

In the case, an ontology is maintained in a remote database system, the application must login to the database and maintain the database connection while a user annotates the contents of object(s). An important issue in the integration of multiple ontologies is handling of heterogeneities in ontology descriptions. For instance, an ontology may be expressed with an ontology definition language such as OML, SOHE, etc., or may be maintained in a database.

A possible solution to the heterogeneity problem is the use a unified ontology markup language that allows integration of multiple ontologies. With this capability, the application may access more than one ontology with a unified view, However, no standard formats for ontology encoding are currently available.

Concepts and their relations are described within <ConceptDS> and </ConceptDS>. Let us consider a concept, say "red Mustang" (see Figure 6) which can be rewritten as "The Mustang is red", More precise description of the concept may be expressed as "A sport car of type Mustang has visual property red". This concept can be expressed with two conceptual elements, "red", and "Mustang" and one semantic relation, "has-color".

Since, in conceptual graph formulation, a link itself represents "has-a" relation, the "has-a-color" relation can be replaced with "color" without losing any intended semantic information. Introduction of dyadic relation "has" in representation provides systematic mapping from natural language to logic. Note that the "name" property may be represented by NameOf, Name-Of or has-name. This ad-hoc definition of relations create unsystematic mapping in semantic encoding. Entity-relation based semantic representation proposed in [2] uses such ad-hoc definitions of semantic relations among media objects.

Our semantic description of a media is based on the Conceptual Graph Formulation (CGF) with strong restrictions in concept type and relation type assignment. CGF is a knowledge representation mechanism in which knowledge is modeled and expressed with concepts and relations. A conceptual graph is a bipartite graph where nodes of one class represent concepts, nodes of the other class represent relations between concepts, and the partially ordered labels represents types and referents [22]. Semantics of CGF is interpreted with first order logic [Gianni Amati and Iadh Ounis, Conceptual Graphs and First Order Logic, The Computer Journal, Volume 43, Number 1, 2000]. One of the reasons for choosing CGF as our semantic annotation mechanism is that it provides a unified way to describe, store and retrieve the contents. In addition, its graphical representation of semantic content is a semi-formal concept description language that strikes a balance between human comprehension and the possibility of computational support.

To achieve the objectives set forth above, we add following rules to the standard CGF notation:

<u>Rule 1.</u> Concept type restriction. A concept type is defined as a full path with semantic relations from an empty (root) concept to a target concept in an ontology.

<u>Rule 2</u>. Relation type restriction. A relation type is defined as a full path with semantic relations from an empty (root) relation type to a target relation type in an ontology.

<u>Rule 3.</u> A referent of a conceptual relation must be predefined in the lexical ontology and it must be a noun.

<u>Rule 4.</u> A referent of a concept must have a concept type.

<u>Rule 5</u>. No general marker "*" (ordinarily available in CGF) is allowed in descriptions.

The first two simply define concept types and relation types. The purpose of the third restriction is to make mapping from natural language to logical expression more systematic. In CFG based content representations, it is imperative to use relation names for clarifying semantic meaning. In fact, we have noticed that annotators often invent their own relation names such as author-of, teacher-of, friend-of in conceptual-graph based knowledge representation. Since those words are fully understood by a binary relation.

The above restrictions ensure that all concept types appearing in descriptions have corresponding conceptual categories defined in ontologies, and all relations can be uniquely resolved by a lexical ontology.

In a CG, a concept is represented as a rectangle and a circle represents relation between concepts. A concept has a concept type and a referent, which is an instance of the concept type. A relation has a relation type and a referent that is an instance of the relation type. Concept types and relation types are managed in a concept hierarchy. Therefore, they provide partial ordering relations. As an example, the following expression is the interpretation of Figure 6, and shows naming features in UCDS.

wordnet:ϕ@entity@object@artifact@instrumentation@transport@motor_vehicle@automobile@car@sport_car("Mustang")∧

wordnet:ϕ@abstraction@attribute@property@visual_property@color@chromatic_color("red")∧
wordnet:ϕ@abstraction@attribute@property@visual_property@color("Mustang","red")

In the above, @ represents the is-a relation, "wordnet:" represents the source of the ontology, and ϕ represents empty concept (the root concept) in the ontology. "wordnet:" emphasizes that the semantic encoding view of the conceptual element being described is based on an ontology called WordNet. A semantic type (concept type in this case), ϕ@entity@object@artifact@instrumentation@transport@motor_vehicle@automobile@car@sport car is a path from the empty concept denoted by ϕ to the target concept "sport_ear". Full path has to be encoded in a description for efficient retrieval and source disambiguation. Users can freely define many referents such as "Mustang", "Avante", or "Ford Explorer". However, a concept type that subsumes such referents must have be defined *a priori* in an ontology. In the above representation, the first two statements describe conceptual elements and the last statement describes the relation between two conceptual elements. Figure 7 illustrates a graphical representation of the concept, "Red Mustang", and its semantic disambiguation process by adding concept type information.

However, single concept type expressed with a term is not enough to handle the issue of polysemy (single term may represents multiple meanings). For instance, a user assigns "horse" (animal horse) as a concept type to a keyword "Joe". However, "horse" has several different semantic meanings, therefore, Joe (animal horse) and "Joe" (a cavalry) may not be successfully discriminated. This is the reason to use full concept path as a concept type description. Another approach proposed by Guranio et al [11], use a synset (a concept in WordNet) as a concept cluster in semantic disambiguation. However, their approach may not efficiently handle and use semantic inheritances and relationships captured by an ontological structure in semantic information retrieval. The reason is that a single synset does not provide direct access to conceptual relations, and thus, additional operations (recursive browsing of conceptual hierarchy) are necessary to process relation-based query processing. In contrast, our approach naturally allows conceptual browsing of concepts, and computation of similarity between either conceptual entities or semantic relations.

As stated before, our semantic description scheme provides clear description of intended meaning through ontology, formal syntax and interpretation of description. An annotation tool such as Media-CAT may generate those descriptions with minimal human interactions.

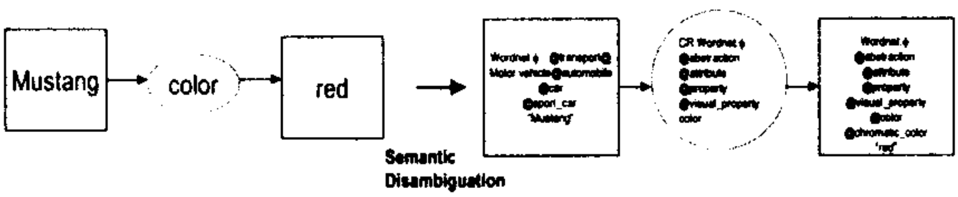

Figure 6: Semantic disambiguation process of a concept "Red Mustang"

A semantic description is a descriptive conceptual graph. The code list shown below is UCDS representation of Figure 6.

```
<SemanticDS>
<ConceptDS ID = "UCDC-21232-3421-321">
    <use-ontology
uri = "http://foo.org/wordnet-ontology/wn-ontology"
use-prefix ="wordnet"
version-info = "1.6">
                <conceptual-element-ds ID = "UCDS-1003-1432DA">
                    <concept-type>

        ϕentity@object@artifact@instrumentation@transport
@motor_vehicle@automobile@car@sport_car
                    </concept-type>
                    <ce-referent>
                        Mustang
                    </ce-referent>
                </conceptual-element-ds>
                <conceptual-element-ds ID = "UCDS-1003-321DA">
                    <concept-type>

        ϕabstraction@attribute@property@visual_property@color@
chromatic_color
                    </concept-type>
                    <ce-referent>
                        red
                    </ce-referent>
                </conceptual-element-ds>
<conceptual-relation-ds>
        <!- describing "media object O1 is a Mustang" ->
                    <relation-type>

        ϕabstraction@relation@linguistic_relation@semantic_relation
                    </relation-type>
                    <cr-referent>
hypernyms
                    </cr-referent>
                    <var pos = 1>
                        <!- Pointing actual syntactic
description of media object O1-->
<MO-DSLink
xmlns:xlink="www.w3.org/2000/xlink"
xlink:href ="#xpointer(id("O1"))">
                        </MO-DSLink>
                    </var>
                    <var pos = 2>
                        <!- Pointing conceptual-element
                description of media object O1-->
                        <CE-DSLink
xmlns:xlink="www.w3.org/2000/xlink"
xlink:href ="#xpointer(id("UCDS-1003-1432DA"))">
                        </CE-DSLink>

                    </var>
                </conceptual-relation-ds>
                <conceptual-relation-ds>
                    <!- describing "the Mustang is red" ->
<relation-type>
        abstraction@attribute@property@visual_property
    </relation-type>
                    <cr-referent>
                        color
                    </cr-referent>
                    <var pos = 1>
                        <CE-DSLink
    xmlns:xlink="www.w3.org/2000/xlink"
    xlink:href ="#xpointer(id("UCDS-1003-1432DA"))">
                        </CE-DSLink>
                    </var>
                    <var pos = 2>
                        <CE-DSLink
    xmlns:xlink="www.w3.org/2000/xlink"
    xlink:href ="#xpointer(id("UCDS-1003-1432DA"))">


                         </CE-DSLink>
                    </var>
                </conceptual-relation-ds>
            </use-ontology>
        </ConceptDS>
    </SemanticDS>
```

UCDS uses a description referencing mechanism provided by XLink [22] and XPointer [23] that is similar to the previous UCDS's media object referencing mechanism presented in [17]. The statement <CE-DSLink> points the actual description of a conceptual element. The need for pointing mechanism is to eliminate the redundancy of descriptions. Similarly, <MO-DSLink> points the actual description of syntactic content description of a media object.

Description structure of the semantic content in UCDS can be seen as two lists. The first list is a conceptual element list, and the second is a conceptual relation list. These two lists are described within <ConceptDS> and </ConceptDS>. User may reuse the existing concepts by using the pointing mechanism. <concept-type> tag is designed to inform the conceptual category of the concept being described. Similarly, <relation-type> describes the semantic category of a relation. <cr-referent> and </ce-referent> are designed to assert user-defined relation literals.

The UCDS's semantic description scheme provides a variety of services for description, classification and retrieval of' multimedia information. A powerful property of our description scheme is the automated determination of subsumption between concepts. Given two concept descriptions C1 and C2, we can determine whether a semantic relation between C1 and C2 is defined as an implicit subset/superset relationship, by comparing the semantic interpretations of C1 and C2 using predicate calculus model theory. It is therefore possible to classify concept descriptions dynamically. A database of concept descriptions can be organized into a partial ordering based on the subset/superset relations. This will provide a taxonomy of database content range from the general to the specific.

Semantic content sharing capability in UCDS is accomplished by using ontology and naming restrictions in a concept type definition and a relation type definition. The naming restriction provides a clear and unique description of a semantic cluster where the semantic description is valid. In addition, semantic relations with other descriptions are also captured.

3.5 Interpretation of Conceptual Graphs in UCDS

UCDS's semantic description represented as a CG requires a knowledge base K, which maintains conceptual type hierarchies. K has three components, $K = \langle T_c, T_r, S \rangle$, where T_c is a set of concept types, T_r, is a set of relation types and S is a set of individual markers. Formally, a descriptive conceptual graph is a quadruple, $G = \langle C, R, E, I \rangle$, where C denotes a set of concepts, R denotes a set of relations, and E denotes edges. I is a mapping a label to a node. Edges incident on a relation are totally ordered. Each concept node has a label given by mapping I. A label for a concept node c_i, has concept-type(c_i) and referent(c_i). Only an individual (instance of a concept or a relation) is allowed in DCG. In other words, referent(c_i) should not be a variable. A relation r_j is labeled by a relation type, relation-type(r_j) and referent(r_j). Given K, and G, the following conditions have to be met.

relation-type(r_j) $\in$ Tr
referent(r_j) $\in$ S
concept-type(c_i) $\in$ Tc

Given *K* and *G*, semantics of first order logic can be used in interpretation of UCDS's semantic annotations. Let ϕ(G) be a formula based on FOL for G. Let G be a CG made up of a binary relation R on the two concept nodes [Cl:x] and [C2:y]. Linear form [sowa] of this G is [Cl:x]->(R:z)->[C2:y]. We denote by CI('x') the translation of [Cl:'x'] in FOL, where 'x' is a referent in CG. If C1 and C2 are the concepts adjacent concept nodes to a relation (R:z), we associate a formula, ϕ(G) as follows:

$\phi(G) = Cl('x') \wedge C2('y') \wedge R('x','y')$

More efficient logic based interpretation of CG is to use type logic syntax that provides more compact representation scheme.

4. Conclusion

In depth of problems and potential solutions for heterogeneities in content description (annotation) is presented in this paper. A problem that annotator and retriever are at the different conceptual spaces (heterogeneity in conceptual view is can be solved by conceptual space description, and semantic heterogeneity can be resolved by explicit description of concept source represented as a concept path in UCDS. The proposed content description scheme is based on the logical foundation and can be used in efficient semantic content browsing and retrieval. Similarity distance measurement between two concepts has already been explored and their efficiency has been reported in [dexa]. A tool for annotating the proposed scheme is in the final testing stage (please check www.mpeg7tv.com).

REFERENCES

1. Gianni Amati and Iadh Onis, Conceptual Graphs and First Order Logic, The Computer Journal, Volume 43, Number 1, 2000

2. B. Benitez, S. Paek, S.-F. Chang, "Object-Based Multimedia Description Schemes and Applications for MPEG-7", To appear in the Image Communications Journal, Invited Paper on a Special Issue on MPEG-7, 2000

3. Dublin, "Dublin Core Metadata Element Set, Version 1.0: Reference Description, http://dublincore.org/documents/1998/09/dces/#

4. Gerard Robert Ellis, Managing Complex Objects, Ph.D. Thesis, Department of Computer Science,
University of Queensland, 1995

5. Christiane Fellbaum (ed.), *WordNet: An Electronic Lexical Database*, MIT Press, 1998

6. D. Fensel, J. Angele, and R. Studer, The Knowledge Acquisition And Representation Language KARL, IEEE
Transactions on Knowledge and Data Engineering, l0(4):527-550, 1998.

7. D. Fensel, J. Angele, S. Decker, M. Erdmann, H-P. Schnurr, S. Staab, R. Studer, and A. Will: On2broker: Semantic-Based Access to Information Sources at the WWW. In Proceedings of the World Conference on the WWW and Internet (WebNet 99), Honolulu, Hawaii, USA, October 25-30, 1999.

8. D. Fensel, M. Crubezy, F. van Harmelen, and M. I. Horrocks: OIL & UPML: A Unifying Framework for the Knowledge Web. In Proc. of the Workshop on Applications of Ontologies and Problem-solving Methods, ECAI'00, Berlin, Germany August 20-25, 2000.

9. M. Flickner et al. "Query by image and video content: The QBIC system", *IEEE Computer*, vol. 28(9), pp23-32, September 1995.

10. N. Guarino, "Concepts, Attributes and Arbitrary Relations: Some Linguistic and Ontological Criteria for Structuring Knowledge Base:, Data and Knowledge Engineering. Vol.8, No2, 1992, Pp. 249-261

11. Nicola Garino et.al., "OntoSeek:Content-based Access to the Web", *IEEE Intelligent Systems*, Vol. 14, No. 3, May/June 1999

12. F. Idris and S. Panchanathan, "Review of Image and Video Indexing Techniques", *Journal of Visual Communication and Image Representation*-Special Issue on Indexing, Storage and Retrieval of Images and Video, June 1997.

13. F. Golshani, Y.C. Park, & S. Panchanathan, "Interactive Indexing and Retrieval of Images and Videos in VideoRoadMap" *SPIE '98 Voice, Video, and Data Communication, Boston* MA.

14. ISO/IEC JTC 1/SC 29/WG 1l/N3815, MPEG-7 Multimedia Description Schemes XM (Version 6.0), January 2001, Pisa, IT

15. Y.C. Park et al, "Polygon based bounding volume as a Spatio-temporal data model for video content access", *Internet Multimedia Management Systems Conference*, Boston, MA, November 2000.
16. Y. C. Park, "Tools for Power Annotation of Visual Content: Lexicographical Approach", *In Proc. of ACM Multimedia* 2000 Nov. 2000, LA, CA, USA
17. Y.C. Park, F. Golshani, & S. Panchanathan, "UCDL: A Language for Audio-visual Content Description", *SPIE '99 Voice, Video, and Data Communication,* Boston MA, 1999.
18. Amit Sheth, Wolfgang Klas eds. Multimedia Data Management using metadata to integrate and apply digital media, McGraw Hill, 1998
19. J. R. Smith and S-F.Chang, "VisualSEEk : a fully automated content-based image query system", *The Forth ACM Multimedia*
20. John F. Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations, PWS Publishing Cc., Pacific Grove, CA, 1999
21. W3C, XML Pointer Language (XPointer) Version 1.0, http://www.w3.org/TR/xptr
22. W3C, XML Linking Language (XLink) Version 1.0, http://www.w3.org/TR/xlink/]
23. Inktomi, UltraSeek, http://www.inktomi.com/products/search/products/ultraseek/cce/ ccetop.htm
24. Software AG, Tamino http://www.softwareag.com/tamino/