Integration of Domain-Specific and Domain-Independent Ontologies for Colonoscopy Video Database Annotation

Jie Bao[1], Yu Cao[2], Wallapak Tavanapong[2], and Vasant Honavar[1]
Artificial Intelligence Research Laboratory[1]
Multimedia Research Laboratory[2]
Department of Computer Science
Iowa State University, Ames, IA USA 50010
{baojie, yucao, tavanapo, honavar}@cs.iastate.edu

Abstract

Flexible and effective indexing of video data so as to render it useful in automated inference and data-driven knowledge acquisition is a key problem in knowledge engineering. In this paper, we present an approach to annotation of a video database using a domain specific ontology, a domain independent video ontology that encodes the structure and attributes of video data. The two ontologies are integrated using domain-specific semantic linkage. The result, an integrated ontology for video annotation (IOVA) is represented in OWL, a description logic based ontology language. We describe a user-friendly platform for video database annotation (VIDAI) using IOVA. We demonstrate the use of IOVA and VIDAI in annotating a database of colonoscopy videos. Video entities are annotated with a combination of domain-independent features and the domain-specific entities. Annotation of video data using IOVA constitutes an important first step towards flexible and fully automated indexing, retrieval, inference, and data-driven knowledge discovery using video data in a broad range of applications including colonoscopy video analysis.

Keywords: *Colonoscopy, MST, OMED, Video Database, Semantic Web, OWL*

1 Introduction

Recent advances in imaging, storage, and communication technologies have resulted in a growing amount of multimedia video data (e.g., videos of medical procedures, news stories, scientific experiments, etc).

For example, colonoscopy is rapidly becoming the single most important endoscopic screening modality for colorectal cancer – which is only second to lung cancer in terms of the number of cancer related deaths. Colonoscopy allows inspection of the entire colon and provides the ability to perform a number of therapeutic operations (e.g., polyp removal) during a single procedure. A sequence of images of the colon generated during endoscopic procedures provides invaluable information for colorectal research. However, these images are not typically captured for real-time or post-procedure review and analysis [11, 12]. To address this need, [3] has developed an experimental system for parsing colonoscopy videos into semantic units, as the first step to establish a content-based analysis system.

A logical next step would be to develop a system for indexing and retrieval of video fragments, for post-procedure review and analysis, and more importantly, data-driven knowledge acquisition for computer-assisted colon cancer diagnosis. Representation of such multimedia video data in a semantically rich, structured form that lends itself to automated search, retrieval, inference, and data-driven knowledge acquisition (e.g., using machine learning) is an important task in knowledge engineering.

Several approaches have been proposed for description of multimedia metadata. Examples include Dublin Core (http://dublincore.org) and RSS (RDF Site Summary) (http://web.resource.org/rss) frameworks for annotation of digital library and online documents and the MPEG-7(http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm) standard for describing multimedia content. Dublin Core and RSS offer essentially a standard set of tags for the annotation with a very limited ability to represent semantic relationships among entities (e.g., creator and image title). MPEG-7 definitions (description schemas and descriptors) are expressed in XML schema. While XML is useful for representation of hierarchically structured name spaces, it is of limited utility as a knowledge representation language [17].

There have been several efforts in specific application domains to develop domain-specific controlled vocabularies. For example, Minimal Standard Terminology (MST) [10] is a controlled vocabulary for colonoscopy and endoscopy reporting. MST was initiated by the European Society of Gastrointestinal Endoscopy (ESGE) and publicized by World Organization of Digestive Endoscopy (OMED). MST lists a core set of terms to be used in this domain. However, MST was initially intended for clinical reporting purposes. Because of the informal nature of MST, MST terms can be semantically ambiguous.

A key challenge is to integrate standard domain independent approaches to annotation of multimedia video content with domain-specific controlled vocabularies or ontologies for specific domains to produce semantically rich annotations. Ideally, such annotations must:

- Describe both the structure of the video as well as domain-specific terms and relationships among terms within a single unified framework.
- Formal (with precise syntax and semantics)
- Representationally adequate – that is, expressive enough to permit precise specification of the natural entity structure of the domain and the relations and constraints among the entities.
- Inferentially adequate – that is, permit automated inference so as to allow automated indexing, search, retrieval, reasoning, and data-driven knowledge acquisition.

Annotation of Video content can be seen as a special case of annotation of web content using a formal and semantically precise language that would allow software agents to *understand*, *deduce*, and *induce* information from annotated web resources. Semantic web [16] is a vision for the future web with enhanced information access based on the exploitation of machine-processable metadata. For example, realization of the semantic web vision would allow an agent to know that *Colonoscopy* is an *Endoscopy* that studies the *Site Colon*. It will enable an agent to infer that if a *Video* is about *Colon* then all scenes in that video are likely to be about *Colon*.

Ontologies facilitate knowledge sharing among users (including software agents), by providing a formal conceptualization of a given domain. In the context of knowledge modeling, an ontology is a description of the concepts and relationships of interest [4]. Consequently, there is a large body of work in artificial intelligence which is focused on development of languages for representing and reasoning with ontologies.

OWL [http://www.w3.org/Tr/owl-ref] is a standard web ontology language supported by the world-wide web consortium. OWL and the associated inference tools offer a powerful knowledge representation framework for video database annotation.

Against this background, this paper describes colonoscopy video annotation using an integrated ontology for video annotation (IOVA), a user-friendly platform for video database annotation using IOVA (VIDAI) and the use of IOVA and VIDAI in annotating a database of colonoscopy videos.

2 Knowledge Modeling

2.1 Colonoscopy Domain Ontology

The colon is a hollow, muscular tube which is about 6 feet long. A normal colon consists of six parts: cecum with appendix, ascending colon, transverse colon, descending colon, sigmoid and rectum. Colonoscopy is a procedure that allows for inspection of the entire colon. A flexible endoscope (a flexible tube with a tiny video camera at the tip) is advanced under direct vision via the anus into the rectum and then gradually into the most proximal part of the colon or the terminal ileum. During a colonoscopic procedure, the tiny analog video camera at the tip of the instrument generates a sequence of frames of the internal mucosa of the colon. These frames are displayed on a monitor. The endoscopist interprets the displayed video and acts based on his/her knowledge regarding the condition of the patient combined with his/her colonoscopic expertise.

The colonoscopy domain is modeled using an ontology based on MST vocabulary for endoscopy reporting. MST is a controlled list of standard terminology in gastrointestinal Endoscopy (GIE) for description of "Reasons for performing the endoscopy", "Findings", "Endoscopic Diagnosis" and other details of examination in GIE report. MST includes 1713 terms: 122 reasons, 8 Endoscopic Procedures, 1030 Findings, 7 Complications, 166 Additional Procedures, 235 Diagnosis, 93 Sites (anatomical), and 52 Details of the examination. However, only a part (26 reasons, 7 Complications, 30 diagnosis, 3 examinations, 38 findings, 15 sites, 8 additional diagnostic procedures) of the terminology is relevant for colonoscopy knowledge modeling (See Figure 1).

The terms and their relations in MST documentation are given in a set of tables. The exact semantic meaning is carefully reviewed since the same table structure is usually used for semantically different settings.

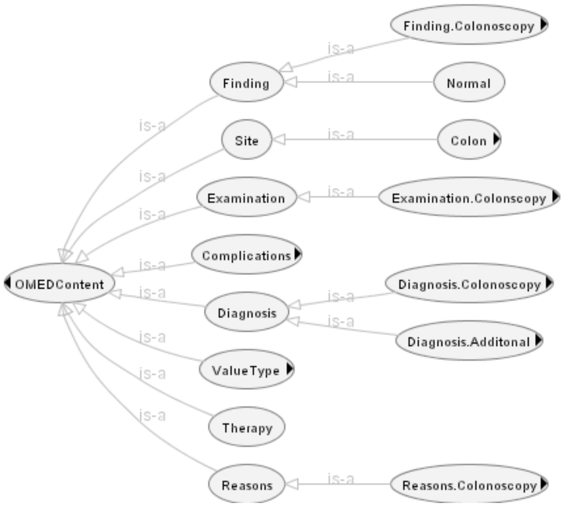

Figure 1: Organization of MST

Class taxonomy

Colonoscopy related MST terms are represented in the form of an ontology modeled using OWL. The original MST table is manually transformed into an OWL-based ontology. All concepts are group as subclass of "OMEDContents" class.

The terms and their relations in MST documentation are given in a set of tables [10]. We start with a class taxonomy of terms, like site, findings, reasons, and diagnosis. Taxonomy is built for each sub categories with terms described in the original tables. However, a lot of terms in the original terminology have duplicated names; for example, "*Polyps*" is used in *Finding*, *Diagnosis* and *Reasons* with slightly different semantic settings. Hence, we disambiguate between semantically distinct uses of names in the MST vocabulary by adding a class hierarchy based path expression as a prefix. For example, the three "polyps" terms are named as

Finding.Colonoscopy.ProtrudingLesions.Polyps
Diagnosis.Colonoscopy.Polyps
Reasons.Colonoscopy.Diseases.Polyps.

Some names with no duplication in colonoscopy terms are also given in this way to ensure the extensibility for the future.

Properties

Terms listed as "Attributes" and "Attributes Values" are modeled as properties of corresponding classes and instances of specially typed value classes when necessary. The level in the class taxonomy where an attribute is placed is based on a tradeoff between compactness and efficiency while guaranteeing sound semantics For example, *hasExtent* is set as a property of *Mucosa* since most of its subclasses have this property; as to the subclasses of *Mucosa* without such property, eg. *Petechiae*. we just set the cardinality of the property to 0. Although our current implementation is limited to the subset of MST that is relevant to Colonoscopy domain, the design can be easily extended to the entire endoscopy domain covered by MST.

2.2 Video Ontology

In this section, we discuss the design of the video ontology.

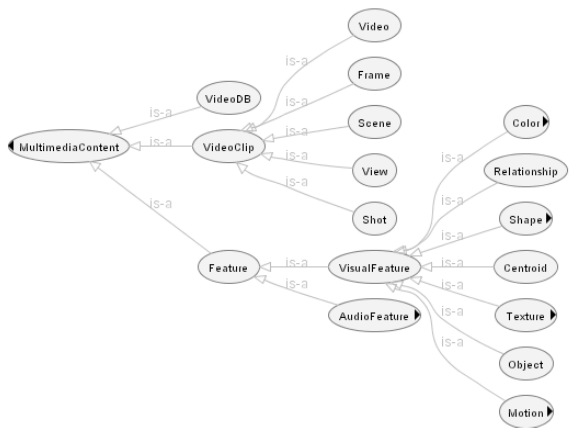

Figure 2: Class taxonomy of video ontology

Class taxonomy

Our video entities are based on two sources: the semantic units derived from the content-based analysis of colonoscopy videos; and MPEG-7 descriptors.

We use an approach to parse the colonoscopy videos into semantic units [3] such as scenes, operation shots, and key-frames. We define a scene as a segment of visual and audio data that correspond to an important part of the colon such as "Rectum", "Sigmoid", and "Descending colon". An operation shot is a segment of visual and audio data that corresponds to a biopsy or therapeutic operation such as polyp removal using a snare. In this paper, we use the term shots to represent operation shots.

MEPG-7 standard does not cover important semantic information in colonoscopy videos. Hence, we had to extend the set of MPEG-7 descriptors e.g., to describe operation shots of the colonoscopy videos.

Figure 2 shows the proposed video entity taxonomy. Note that the top-level class "*MultimediaContent*" is the superclass of all concepts related to video descriptors, such as VideoClip, VideoDB, and

Feature. The reason is that the hierarchical structure of a video is modeled using partOf relations as opposed to subClassOf relations. The use of subClassOf (or ISA relationship) is inappropriate in this case because lower level class should not inherit properties of high level classes. For example, the key-frame of a scene may not be the same key-frame for its component shots, and automatic heritance of properties is harmful in this case. However, the video segment classes also share some common properties like *hasSite*, and *hasID*. All those common properties are modeled in a common interface called *VideoClip* and all the video segment classes are derived from it.

Properties and restrictions

All partOf relations are implemented by *hasStart*, and *hasEnd* properties. Their domain and range constraints are different at each segment level. For example, the range of hasStart and hasEnd of class *View* is restricted to *Scene* while that of *Scene* is restricted to *Shot.*

```
<owl:Class rdf:ID="View">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#VideoClip"/>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:FunctionalProperty
rdf:about="#hasStart"/>
        </owl:onProperty>
        <owl:allValuesFrom>
          <owl:Class rdf:about="#Scene"/>
        </owl:allValuesFrom>
      </owl:Restriction>
    </rdfs:subClassOf>
……
  </owl:Class>
<owl:Class rdf:ID="Scene">
    <rdfs:subClassOf rdf:resource="#VideoClip"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:FunctionalProperty
rdf:about="#hasStart"/>
        </owl:onProperty>
        <owl:allValuesFrom rdf:resource="#Shot"/>
      </owl:Restriction>
    </rdfs:subClassOf>
……
  </owl:Class>
```

Figure 3: Definition of class Scene and View

A part of the OWL source is shown in Figure 3

In OWL, there are two types of properties: "Datatype Property" and "Object Property". The "Datatype Property" represents the simple property of the entity, like "shot ID" and "shot annotation" for entity "*Shot*", The object property implies a relationship between two entities. We use a function, with the current entity as its domain and another entity as its range, to represent this property (relationship). Typical relationships are "part of", and "subclass of". Take the entities "*VideoClip*" and its subclass "*Video, View, Scene, Shot*", and "*Frame*" as an example, Figure 4 illustrates the relationships among those entities.

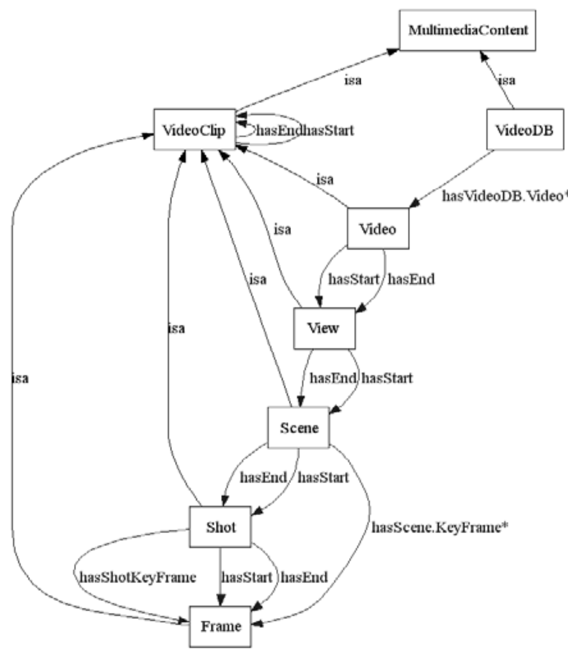

Figure 4: The Properties and Relationships among the "VideoClip" and its subclasses

2.3 Integration of Domain Ontology and Video Ontology

The two ontology components, video ontology and MST domain ontology, are integrated to ensure meaningful annotation of video data so as to allow indexing, retrieval, and eventually, data-driven knowledge discovery. Figure 5 shows the relationships between the two ontologies.

Instances of MST ontology are added as features to the instances of video ontology. This allows us to annotate different levels of video ontology with different MST annotations. For example, the *Reasons* and *Examination* properties are only meaningful at the *Video* level, but not at the *Frame* or *Shot* levels. However, some properties can appear at multiple structural levels. For example, Site could be used at both *Video* level (such as *Colon*) and shot level (such as *Colon.Transverse*). We associate properties that appear at multiple levels with the *VideoClip* interface. Details of the annotation level of MST terms on the video ontology are shown in Table 1.

Cardinalities of those MST are all set to >=1, which means multiple annotation of them are legal. Since all subclasses of those MST basic classes can also be processed as their ancestor classes, polymorphyism is enabled. For example, instances of *Symptoms.WeightLoss* and *Symptoms.Diarrhea* can be added as some *Reasons* properties of a given *Video* instance even those two classes are not stated in the video properties.

Video Level	**Associated basic MST classes**	**Cardinality**
Video	Reasons	>=0
	Examination	>=0
	Therapy	>=0
VideoClip(inherited by all)	Site	>=0
	Finding	>=0
	Diagnosis	>=0

Table 1: annotation level of MST terms on video ontology

A simplified sample instance built on this ontology is given in figure 6. The ontologies shown were created using the Protégé ontology editor[1], and visualized using the OntoViz and OWLViz plugins.

3 Related work

3.1 Video ontology design for semantic web

Recent years have seen some research on multimedia semantic web. J. Hunter [7] introduced the techniques to add multimedia to semantic web by building MPEG-7 ontology in DAML+OIL language. However, the proposed techniques can not be used to represent all the semantic units for colonoscopy videos since the ontology in that paper is originated from MPEG-7 without any extension. Grosky et al. presented techniques to build multimedia semantic web for a collection of linked multimedia documents such as arts collection [6]. The core idea is to capture and represent the semantics of the documents by the identification of various users' browsing paths through those multimedia collections. This approach is not suitable for colonoscopy videos because we are unable to build the domain ontology based on the users' behavior.

Troncy [13] discussed a system that integrates both structure and semantics of audio-visual documents. The overall architecture includes a OWL/RDF for a domain specific ontology, a XML-Schema based MPEG-7 ontology, and the transformation mechanisms between the two ontologies. Guler and Pushee [5] introduced a framework for mining video data. The framework consists of both video description schema and intuitive browsing. The proposed description schema is based on the structure and the semantics of the video data. The schema also incorporates scene, camera, and object information pertaining to a large class of video data. The main limitation of this system is that the relational schema used in this system is not as expressive as the ontology language. Addis *e. al.* [1] introduced SCULPTEUR, an architecture for integrated concept and metadata for content based browsing and retrieval for museum applications

3.2 Domain specific ontologies

There is a large body of work on the construction of domain specific ontologies in many areas. Examples of such ontologies in the endoscopy domain include SNOMED DICOM microglossary [8] and UMLS [14]. Strategies for mapping strings in gastrointestinal endoscopy reports to UMLS meta-thesaurus based on natural language processing methods have been studied [19]. HealthCyberMap [2] models a subset of Dublin Core metadata for general health information. The Dublin Core subject field is populated with UMLS terms imported directly from UMLS knowledge source servers. The conversion from UMLS Semantic Network into OWL has also been studied [9]. However, because of the ambiguous syntax and semantics of semi-formal representations, those approaches are not semantically robust to build ontology on MST.

4 Conclusions and future work

In this paper, we have described our experience with building an integrated ontology of colonoscopy video. This work constitutes an important first step towards flexible and fully automated indexing, retrieval, inference, and data-driven knowledge discovery using video data in a broad range of applications including colonoscopy video analysis. Future work includes the design and implementation of query processing engine based on the proposed ontology, a user-friendly ontology editor and annotator, and appropriate performance measurements for performance comparison between our proposed system and other similar systems.

[1] http://protege.stanford.edu/

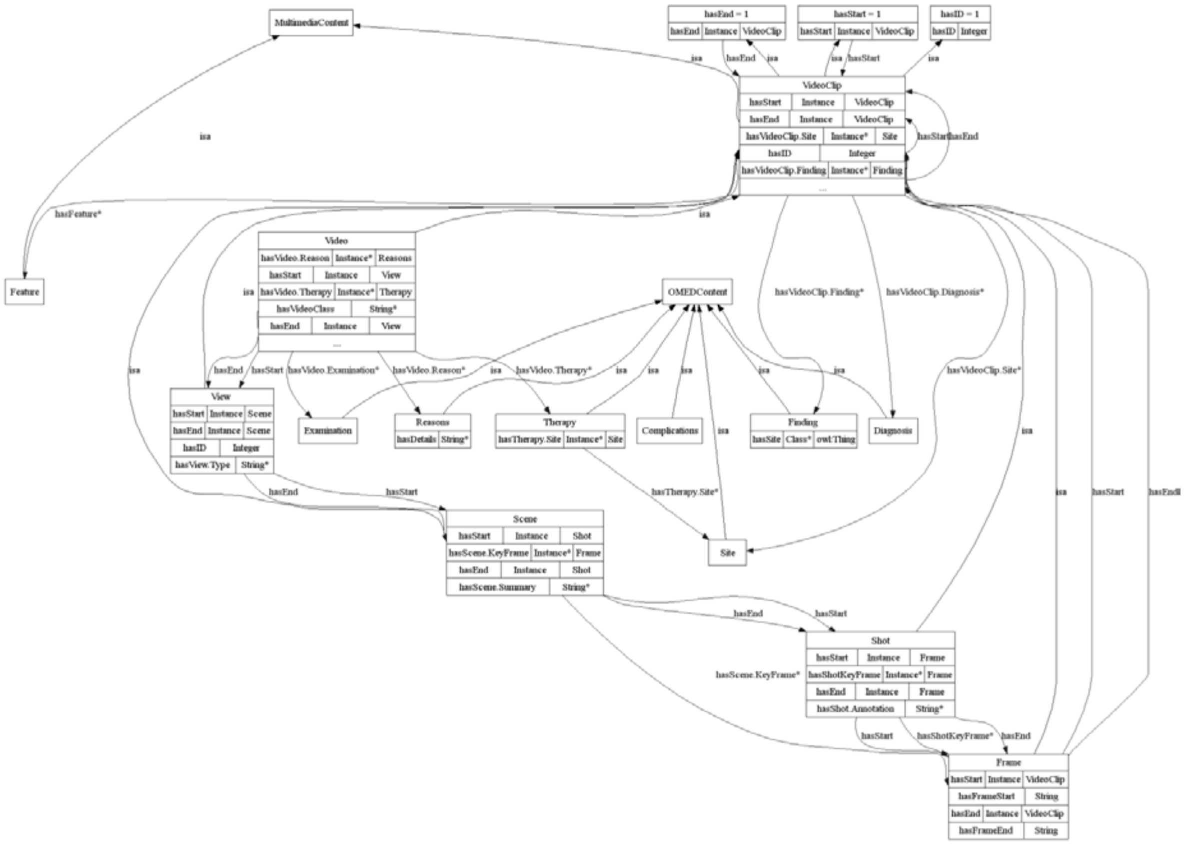

Figure 5: the Integration of multimedia ontology and the MST domain ontology

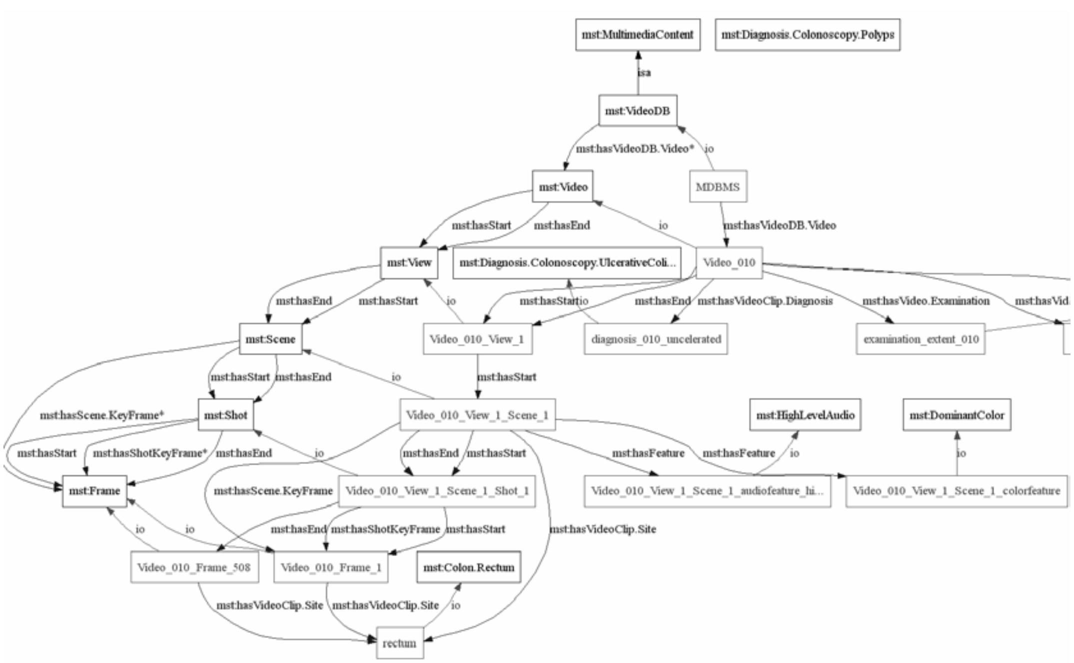

Figure 6: A sample instance of this ontology

Acknowledgements

This research is supported in part by grants from the National Science Foundation to Vasant Honavar (0219699) and Wallapak Tavanapong (0092914).

References

1 Addis, M. Boniface, M., Goodall, S., Grimwood, P.. SCULPTEUR: Towards a New Paradigm for Multimedia Museum Information Handling. In the Proceeding of ISWC2003, D. Fensel et al. INCS 2870, pp. 582-596, 2003.

2 Boulos, K., Roudsari, M., Carson E. (2002). Towards a semantic medical Web: HealthCyberMap's tool for building an RDF metadata base of health information resources based on the Qualified DublinCore Metadata Set. Med Sci Monit,2002;8(7):MT124-136

3 Cao, Y., Tavanapong, W., Kim, K., Oh, J-H., de Groen, P. A Framework for Parsing Colonoscopy Videos for Semantic Units, to Appear in Proceedings of the 2004 IEEE International Conference on Multimedia and Expo, Taipei, Taiwan, June 27 ¨ July 30, 2004

4 Gruber, T. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220, 1993.

5 Guler, S., Pushee. I. Videoviews: A Content Based Video Description Schema and Databse Navigation Tool. In Mining Multimedia and Complex Data, LNAI 2797, O.R.Zaiane et al.(eds.) pp. 134-148, 2003.

6 Grosky, W., Sreenath, D.V. Fotouhi, F. "Emergent Semantics and the Multimedia Semantic Web" in SIGMOD Record, Number 4, December 2002

7 Hunter, J. Adding Multimedia to the Semantic Web – Building an MPEG-7 Ontology. In the proceeding of the First International Semantic Web Working Symposium (SWWS'01), p.261-283, Stanford, California, 2001

8 Korman L.Y. Bidgood W.D Jr. Presentation of the Gastrointestinal Endoscopy Minimal Standard Terminology in the SNOMED DICOM microglossary. Proc. AMIA Annu Fall Symp. 1997:434-8.

9 Kashyap, V. and Borgida, A. Representation the UMLS Semantic Network Using OWL, In the Proceeding of ISWC2003, D. Fensel et al. INCS 2870, pp. 1-16, 2003.

10 Minimal Standard Terminology Digestive Endoscopy: International edtion, April 22, 1998, http://www.omed.org/minimal.htm. Accessed on March 14, 2004

11 Padgett, M. and Sibbett, W. "An Endoscopic Imaging System for the early detection of cancer." (final report for EPSRC grant) http://www.physics.gla.ac.uk/Optics/projects/cancerFluorescenceImaging/Endoscope1.pdf

12 L. E. Sucar and D. F. Gillies, "Knowledge-based assistant for colonoscopy," in Proc. of the 3rd Int'l Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, June 1990, vol. 2,

13 Troncy, R. Integration Structure and Semantics into Audio-visual Documents. In the Proceeding of ISWC2003, D. Fensel et al. INCS 2870, pp. 566-581, 2003.

14 Tringali, M., Hole, W..T. Srinivasan, S. Integration of a Standard Gastrointestinal Endoscopy Terminology in the UMLS Metathesaurus. In the Proceeding of the AMIA 2002 Annual Symposium, Page 801.

15 Tringali, M., Rindflesch, R. Kilicoglu, H. Fiszman, M. and Bodenreider, O. Strategies for mapping concepts in gastrointestinal endoscopy reports to the UMLS (2004).

16 Berners-Lee, T., Hendler, J., and Lassila, O. The Semantic Web, Scientific American, May 2001.

17 Klein, M., XML, RDF, and Relatives (short tutorial). IEEE Intelligent Systems, special issue on "Semantic Web Technology", 16(2):26-28, March/April 2001