

MPEG-7 as a Metadata Standard for Indexing of Surgery Videos in Medical E-Learning

Andrzej A. Kononowicz and Zdzisław Wiśniowski

Department of Bioinformatics and Telemedicine,
Jagiellonian University, Medical College,
ul. Kopernika 7e, 31-034 Krakow, Poland
a.kononowicz@cyfronet.pl

Abstract. The analysis of video recorded surgical procedures is considered to be a useful extension of the medical curriculum. We can foster the development of video-based e-learning courses by working out a unified description method which would facilitate the exchange of these materials between different platforms. Sophisticated metadata enables a broader integration of artificial intelligence techniques into e-learning. The aim of this paper is to present the possibility of combining the MPEG-7 metadata standard with the MeSH classification for indexing of video recordings in medical e-learning. A tool for metadata descriptions of surgical videos in accordance with the MPEG-7 standard is also presented. This tool is part of a larger architecture for the exchange of medical multimedia objects.

Keywords: MPEG-7, e-learning, metadata, medical terminology, learning objects, MeSH.

1 Introduction

E-learning is a contemporary way of learning using multimedia and communication abilities of modern computers and mobile devices (cellular phones, smartphones, palmtops). The main obstacle in the popularization of e-learning is the significant expenditure of time and costs to prepare multimedia materials. We may reduce the problem by building large databases of educational components, which can be used in building more complex courses. Such elements are called learning objects or sharable content objects [1]. On one hand a collective creation and usage of didactic components ensures a larger choice of materials, on the other hand it allows to share the development costs. An example of an initiative promoting e-learning through exchange of learning objects in medicine is the eViP project [2][3]. In this paper we will concentrate on a selected subclass of learning objects which are multimedia learning objects designed to be used in medicine teaching.

It seems to be obvious that the illustration of traditional textual descriptions of procedures by video clips provides the students with broad knowledge. Multimedia learning objects consist of audio/video recordings created during selected surgical procedures. Multimedia databases containing learning objects are a very valuable

source of data for medical educators, however only under the condition that the materials are easily searchable. For that reason, an appropriate description of the content is a crucial part of every multimedia learning object. The description should be easily processed and portable. Metadata designated for medical e-learning should also fulfil some additional requirements specific to the field of medicine – e.g. embedding of existing medical terminologies.

1.1 Usage Scenarios of Multimedia Metadata in Medical E-Learning

Let us consider the case of a student, who is looking for video recordings of minimally invasive surgical procedures. The student enters the keyword “endoscopy” into the search engine. In response, the search engine contacts (for instance by a software agent) a terminology server linked with a medical ontology to look up the definition of the term “endoscopy”. It gets the information that endoscopy is a surgical procedure which has several subclasses – e.g. Arthroscopy, Colonoscopy, Gastroscopy or Thoracoscopy. Additionally, the agent contacts also a user preferences database to get the profile of the user (preferred video format, connection bandwidth, accepted languages). Such databases can be integrated with the university’s learning management system. The user database may also contain certificates which entitle the student to enter educational database with restricted access. Within the next step, the search agent queries the multimedia databases using the information obtained from the terminology server, web ontologies and the user preferences database. Results obtained from many databases are unified and presented to the user, who selects the adequate material. The list of potential search criteria is large – it can include affected organs or used equipment. Due to a spatial and temporal decomposition of the video file, the transmission can start from the right time point and with the desired organ or equipment highlighted.

Another potential use case example of multimedia metadata is to aid the construction of virtual patients [4]. A medical teacher may use a special authoring tool for building complex patient scenarios (e.g. designated for the OSCE examination). The program analyses the data provided by the teacher and automatically suggests videos and images which could be inserted into the virtual patient. The search algorithms are based on metadata stored in multimedia databases.

1.2 Architecture of a Multimedia E-Learning System in Medicine

From the above described use cases we can derive a theoretical e-learning system architecture exploiting multimedia metadata (Fig.1). Surgery videos are stored in media storage systems and transmitted to users by video streaming servers (e.g. Darwin, Helix or Windows Media Services). Video clips are described by physicians or technical staff trained in indexing of medical multimedia resources using specialized annotation tools. The created metadata is stored in a multimedia database, whereas the medical knowledge is kept on a terminology server (there are already free terminology servers available on the Internet – e.g. UMLS) and as web ontologies (e.g. in OWL format). The user profiles and authorization data are integrated with e-learning platforms by system administrators operating dedicated administrative applications and scripts. Learning objects gathered in the system are used by students

(using an interactive learning environment via a standard web browser) or medical educators (using e-learning authoring tools). This paper will concentrate on a selected element of this architecture – the annotation tool for teachers. We will also consider the question what metadata standard should be used.

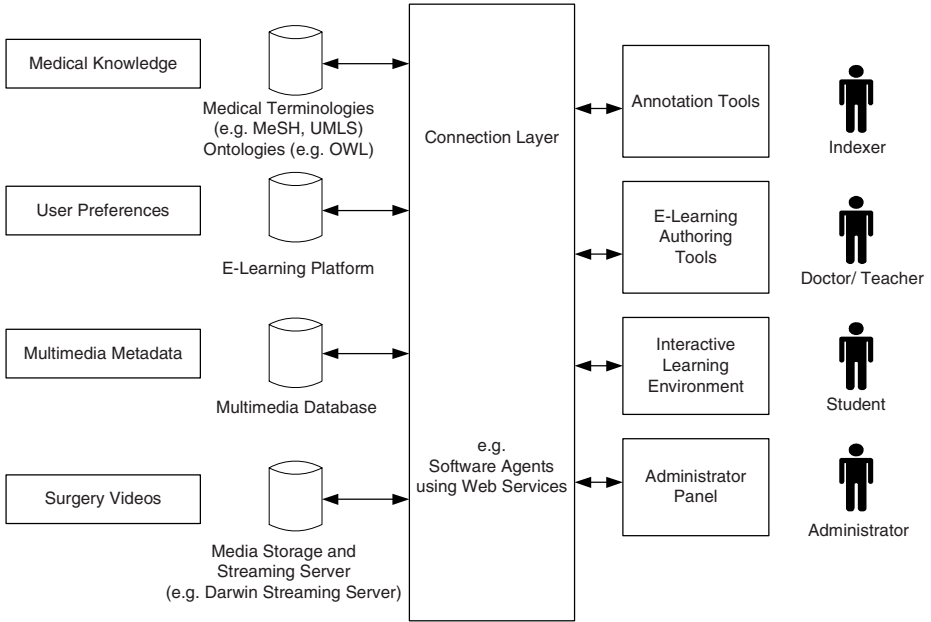


Fig. 1. Architecture of a multimedia e-learning system in medicine

2 Metadata Standards and Terminologies

Much research has already been done in the field of multimedia data description in the past few years. The proposed solutions were based on project specific meta-data, such standards as SMIL, Dublin Core, SMPTE, EBU, TVAnytime or other technologies connected for instance with the Semantic Web initiative (RDF, RDF Schema, OWL, RuleML) [5-8].

2.1 MPEG-7

A significant breakthrough in this field was the release of the MPEG-7 standard in 2001 [9-11]. The official name of the MPEG-7 standard is ISO 15938 Multimedia Description Framework. MPEG-7 is a very flexible specification. It is organized as a collection of tools, which can be used in accordance with the needs of the indexer. The basic building blocks of MPEG-7 are descriptors. Descriptors represent the syntax and semantical meaning of basic description elements (e.g. author’s name, media duration, used codecs, textures, audio signal parameters, camera motion descriptions). Descriptor Schemes (DS) consist of related descriptors and smaller

description schemes. The syntax of MPEG-7 is based on XML. Descriptors and description schemes are defined in the Description Definition Language (DDL), which is an extension of the XML Schema. DDL allows the user to add new elements to the description. MPEG-7 can be stored and transmitted in textual format or as binary data added to a video stream.

The MPEG-7 metadata descriptors have already been prototypically implemented by several research teams. Tseng et al. [12-13] developed a personalization and summarization system consisting of a MPEG-7 video annotation tool, automatic labeling tools and a summarization and adaptation engine. Caliph and Emir [14] are two applications facilitating semantic descriptions of digital photographs. Tsinaraki et al. [15] proposed a video segmentation tool conform to MPEG-7 for ontology-based semantic indexing. Their tool has been tested in the domain of soccer games. Despite the outstanding possibilities of MPEG-7, there is still lack of applications exploiting this standard in medicine. The authors believe that MPEG-7 as a universal, easily extensible and complex metadata standard is the right choice for the description of learning objects also in the medical domain.

2.2 Terminology – MeSH

The diversity of the natural language hinders the automatic processing of descriptions. Therefore, natural language is often artificially limited to concepts stemming from controlled vocabularies (like classification systems or nomenclatures). Many classification systems and nomenclatures have been created to describe medical knowledge (e.g. ICD, LOINC, Snomed CT, NANDA or MeSH). We have decided to use the MeSH (Medical Subject Headings) thesaurus [16] for the description of medical videos in MPEG-7 standard. This vocabulary has been created by the National Library of Medicine (NLM) with the intention to classify information in the biomedical area. MeSH is successfully used in indexing the MEDLINE database and the NLM catalogues. Different language versions of MeSH (e.g. English, German, French or Polish) already exist. The 2006's version of MeSH contains 23885 descriptors. MeSH descriptors are the building blocks of this classification. The elements are divided into 16 categories (e.g. A: Anatomic, C: General Diagnosis or D: General Drugs and Chemicals), which are divided into further subclasses. Descriptors include a set of semantically related concepts, which consist of one or more terms. A descriptor can have attached attributes, which come from a set of 86 qualifiers (e.g. *abnormalities*, *injuries* or *statistics&numerical data*). Hierarchical (*narrower term/broader term*) and non-hierarchical (*related/see also*) relations exist between the descriptors.

Medical terminologies like MeSH are good starting points for semantic description providing the user with a static knowledge reference. If a more advanced semantic search is needed, additionally, the use of ontologies should be considered. Ontologies encode meanings separately from application code enabling knowledge sharing and support for external reasoning. There exist already many examples of ontologies in the medical domain modeling patient data as well as diagnostic and treatment procedures (e.g. [17]). Jovic et al. [18] explain in their study the construction process of medical ontologies on the example of the heart failure domain. They emphasize the importance of the linkage between ontologies and terminologies. The leading language for

expressing ontologies is currently OWL (Web Ontology Language). OWL is usually written in XML/RDF syntax and can be extended by the SWRL rules language. Tsinaraki et al. [19] proposed a framework, called DS-MIRF, for the integration of OWL ontologies with MPEG-7 compliant indexing.

3 MPEG-7 in the Description of Medical Learning Objects

3.1 Video Decomposition

The description of a video file in MPEG-7 may refer not only to the whole clip but also to its fragments. A spatial or temporal decomposition may be distinguished. The first one, expressed by *SpatialDecomposition DS*, allows selecting segments of a picture (e.g. pathological changes, applied medical equipment) and inserting descriptions only of the selected parts. The temporal decomposition (*TemporalDecomposition DS*) enables the partition of the clip into time intervals. It gives the possibility to describe the individual stages of the procedure separately (e.g. preoperative operation, incision, main part of the operation, laying sutures).

3.2 Medical Classifications in MPEG-7

The MPEG-7 standard enables the definition of new classification schemes or importing of the existing ones. The definition of new classifications in MPEG-7 is carried out by the description schemes *ClassificationScheme DS* and *TermDefinition DS*. Concepts derived from declared classifications are placed into the description by

Table 1. Mapping of the MPEG-7 structural annotation classes onto categories of MeSH descriptors

MPEG-7 Structured Annotation	Description of surgery videos	MeSH Categories, Subcategories
Why	Reason for carrying out the procedure. Patient's diagnosis.	C (General Diseases), F3(Mental Disorders)
WhatObject	Names of operated organs or those organs which are visible in the video and are important for students in the opinion of the medical educator.	A (General Anatomy)
WhatAction	Names of performed procedures.	E1-E6 (General Techniques)
How	Medical equipment used in the procedure.	E7 (Equipment and Supplies)
Where	Name of the geographic region in which has the procedure been made.	Z (Geographic Locations)
When	- / No mapping.	-
Who	Patient's characteristic. Patient's profession.	M (Persons)

elements of type *TermUse* and *ControlledTermUse*. This paper focuses on the binding of MeSH terms into the structural description of surgical videos. The structural description in MPEG-7 is represented by the type *StructuredAnnotation*. It may contain any number of *TermUse* instances from seven different categories: actions (element *WhatAction*), persons (*Who*), objects (*WhatObject*), places (*Where*), time (*When*), purpose (*Why*). Each element can enclose a free text description or a reference to a concept from a classification. Table 1 contains the authors' proposal of mapping the MeSH categories onto the MPEG-7 structured annotation types. We also suggest the possible use of the MPEG-7 annotation categories in description of surgical procedures. For instance the element of type *Why* should be used in the description of patient's disease diagnosed, which was the reason for carrying out the operation. We can describe this category by descriptors from the C (*General Diseases*) and F3 (*Mental Disorders*) MeSH subtree.

Example of MPEG-7 code containing a MeSH descriptor.

```
<ClassificationAlias alias="mesh"
  href="http://www.ncbi.nlm.nih.gov/mesh" />
<!-- ... -->
<TextAnnotation>
  <StructuredAnnotation>
    <WhatAction href=":mesh:D013906">
      <Name xml:lang="en">Thoracoscopy</Name>
    </WhatAction>
    <WhatObject href=":mesh:D008168">
      <Name xml:lang="en"> Lung </Name>
    </WhatObject>
  </StructuredAnnotation>
</TextAnnotation>
```

It should be mentioned that the MeSH thesaurus already contains a special qualifier grouping elements for the description of surgical procedures – SU/surg. The qualifier comprises of the following categories: A1-5, A7-10, A13, A14, A17, C1-23, F3. This qualifier could be used theoretically to pick out the MeSH descriptors needed for describing surgical videos. However, in our opinion, the selected set of descriptors is too narrow to fit all concepts useful in the characterization of surgical videos. For instance the qualifier SU/surg does not contain the subcategory E7 – Equipment and Supplies.

4 M7MeDe

As the first element of the proposed architecture of multimedia e-learning systems in medicine, we have decided to implement an annotation tool for surgical videos, which creates descriptions in the MPEG-7 standard using the MeSH classification (Fig.2). The application has been named M7MeDe.

M7MeDe is designed to support the indexation of resources in a medical multimedia library of surgical video recordings. The application enables a temporal

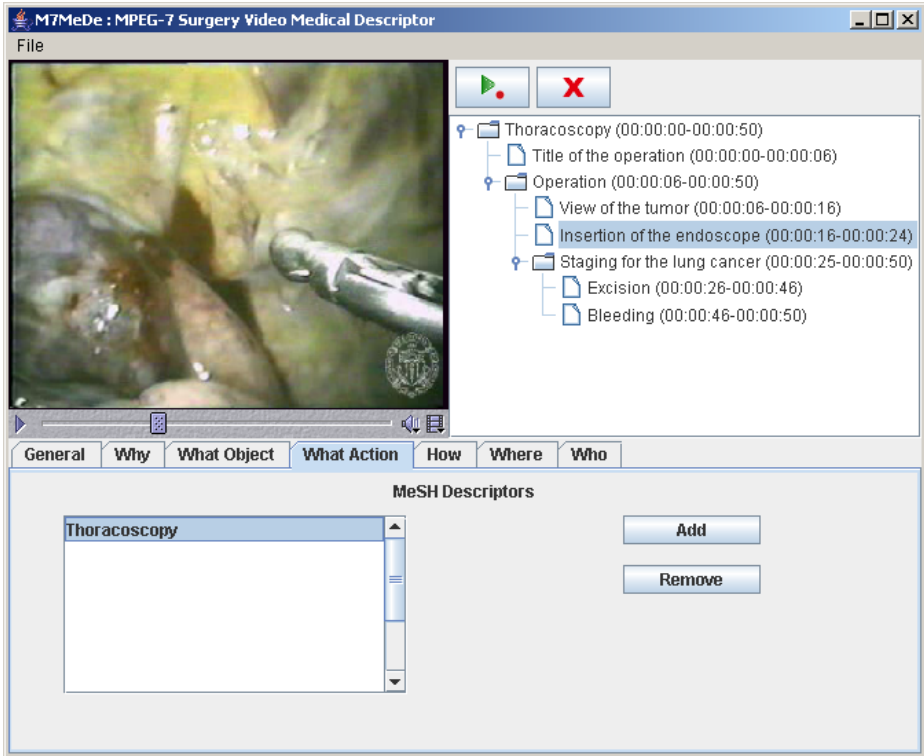


Fig. 2. M7MeDe – an application for describing of multimedia learning objects in the MPEG-7 standard

decomposition of the video, which can be nested in larger segments forming a hierarchical structure. An example of such decomposition carried out with the use of M7MeDe is depicted in Fig 2. The annotated video clip (source of the video [20]) presents a fragment of a thoracoscopy procedure for staging of lung cancer. The first level of video clip description is divided into two parts: the operation’s title screen (displaying the operation’s title and surgeon’s name) and the main operation part. The second part is divided further into three parts: view of the tumor, insertion of endoscope and sampling. The sampling part contains an *excision* and *bleeding* video fragment. We can describe each segment in free text (the *General* tab in Fig 2) or attach keywords from the MeSH thesaurus to a selected MPEG-7 structured annotation category (*Why*, *What Object*, *What Action*, etc). Each annotation category is linked to a subtree in the MeSH-Tree in accordance with the mapping in Table 1. For instance adding a keyword in the *Why* category opens the *General Disease* subtree (Fig.3). Keywords are inherited by subordinated segments (segments which are nested in other segments).

M7MeDe was implemented in Java technology. The MeSH classification was downloaded in the form of XML files, transformed and inserted into a relational

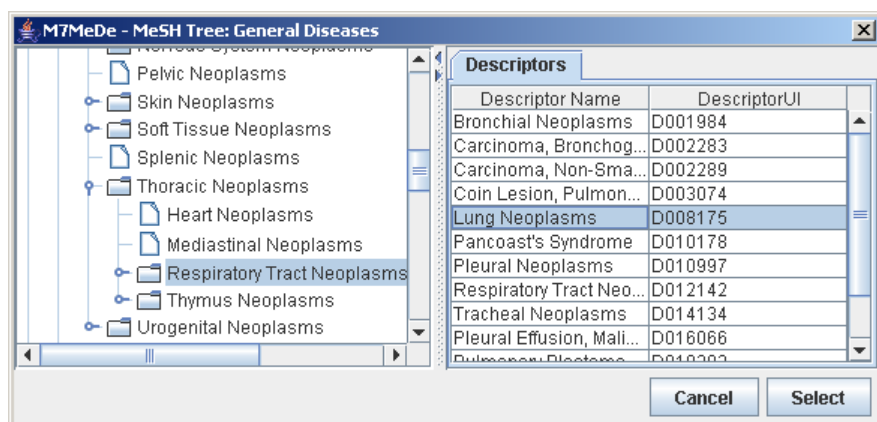


Fig. 3. M7MeDe – Window for selecting of MeSH descriptors in the *General Diseases* category

database. The application uses the Java Media Framework (for video operations) and JAXB (for Java to XML binding).

5 Further Work

The M7MeDe application is still in development stage. Many potential functions are missing (for instance the spatial decomposition of the video or a direct interface to multimedia databases). It is intended to examine the possible relations of medical learning objects described in MPEG-7 to the e-learning standards SCROM and IMS in the future. Further work on this project should also pertain to the remaining elements of the proposed architecture (ontology based search algorithms, placing of multimedia learning objects into learning management systems or construction of authoring tools). Woods et al. [21] showed in their study about indexing of dermatology images, that the use of MeSH alone for indexing finds matching for about one-fourth of the terms in their experiment, therefore usage of other classifications and ontologies in the MPEG-7 description beside MeSH will be necessary. For that reason we consider building direct interfaces to the UMLS terminology server and to repositories of web ontologies in the OWL standard.

6 Summary

Well described surgical video recordings are considered to be valuable e-learning materials in medicine teaching. This paper was aimed at discussing the possibilities of using MPEG-7 and MeSH in building multimedia learning objects. The presented application – M7MeDe – allows the division of a video clip into temporal segments and their description with MeSH keywords. The M7MeDe annotation tool is part of a

larger architecture which takes advantage of medical metadata. There are many possible ways of extending the presented tool by further features.

References

1. Kononowicz, A.A., Żabińska, M.: Distribution of Learning Objects Based on Agents Technology. *Automatyka* 9(1-2), 115–126 (2005)
2. eViP – Electronic Virtual Patients Project, <http://www.virtualpatients.eu>
3. Kononowicz, A.A., Stachoń, A.J., Roterman-Konieczna, I.: Virtual Patient as a Tool for Problem Based-Learning in the Context of the European Project eViP. *E-mentor* 1(23), 26–30 (2008) (in Polish)
4. Huang, G., Reynolds, R., Candler, C.: Virtual Patient Simulation at U.S. and Canadian Medical Schools. *Academic Medicine* 82(5), 446–451 (2007)
5. Mrozowski, P., Kononowicz, A.A.: DSS-MEDA – A Web-Based Framework for Video Annotation in Medical E-Learning. *Bio-Algorithms and Med-Systems* 4(2), 51–56 (2006)
6. Hunter, J., Iannella, R.: The Application of Metadata Standards to Video Indexing. In: *Proc. of the 2nd Europ. Conf. on Research and Advanced Technology for Digital Libraries*, pp. 135–156. Springer, London (1998)
7. Stamou, G., van Ossenbruggen, J., Pan, J., Schreiber, G.: Multimedia Annotations on the Semantic Web. *IEEE MultiMedia* 13(1), 86–90 (2006)
8. Carro, S., Scharcanski, J.: A Framework for Medical Visual Information Exchange on the WEB. *Comp. Biol. Med.* 36, 327–338 (2006)
9. Martínez, J., Koenen, R., Pereira, F.: MPEG-7: The Generic Multimedia Content Description Standard, Part 1. *IEEE MultiMedia* 9(2), 78–87 (2002)
10. Martínez, J.: MPEG-7: Overview of MPEG-7 Description Tools, Part 2. *IEEE MultiMedia* 9(3), 83–93 (2002)
11. ISO/MPEG N4242, Text of ISO/IEC Final Draft International Standard 15938-5 Information Technology - Multimedia Content Description Interface – Part 5 Multimedia Description Schemes, MPEG Multimedia Description Schemas Group, Sydney (2001)
12. Lin, C.-Y., Tseng, B.L., Smith, J.R.: VideoAnnEx: IBM MPEG-7 Annotation Tool for Multimedia Indexing and Concept Learning. In: *Proc. IEEE Intl. Conf. on Multimedia and Expo (ICME)*, Baltimore, MD (2003)
13. Tseng, B.L., Lin, C.-Y., Smith, J.R.: Using MPEG-7 and MPEG-21 for Personalizing Video. *IEEE MultiMedia* 11(1), 42–52 (2004)
14. Lux, M., Klieber, W., Granitzer, M.: Caliph & Emir: Semantics in Multimedia Retrieval and Annotation. In: *Proc. of 19th CODATA Conference*, Berlin, Germany (2004)
15. Tsinarakis, C., Polydoros, P., Kazasis, F., Christodoulakis, S.: Ontology-based Semantic Indexing for MPEG-7 and TV-Anytime Audiovisual Content. *Special Issue of Multimedia Tools and Application Journal on Video Segmentation for Semantic Annotation and Transcoding* 26, 299–325 (2005)
16. Nelson, S., Johnston, D., Humphreys, N.: Relationships in Medical Subject Headings. In: *Bean, C., Green, R. (eds.) Relationships in the Organization of Knowledge*, pp. 171–185. Kluwer Academic Publishers, New York (2001)
17. The National Center for Biomedical Ontology, <http://bioontology.org>

18. Jovic, A., Prcela, M., Gamberger, D.: Ontologies in Medical Knowledge Presentation. In: Proc. 29th International Conference Information Technology Interfaces, Cavtat, Croatia (2007)
19. Tsinaraki, C., Polydoros, P., Christodoulakis, S.: Integration of OWL Ontologies in MPEG-7 and TV-Anytime Compliant Semantic Indexing. In: Persson, A., Stirna, J. (eds.) CAiSE 2004. LNCS, vol. 3084, pp. 398–413. Springer, Heidelberg (2004)
20. Video Assisted Thoracoscopy, University of Southern California, <http://www.cts.usc.edu/videos-mpeg-vidassistthoracoscopy-all.html>
21. Woods, J., Sneiderman, C., Hameed, K., Ackerman, M., Hatton, C.: Using UMLS metathesaurus concepts to describe medical images: dermatology vocabulary. *Comp. Biol. Med.* 36, 89–100 (2006)