SAAMAR: Semantic Annotation Architecture for Accessible Multimedia Resources

Fernando Paniagua Martín, Ángel García Crespo, Ricardo Colomo Palacios*, Belén Ruiz Mezcua

ABSTRACT

The semantic annotation (SA) of resources in general, and particularly of multimedia resources, is an arduous task which developments in automatic annotation mechanisms have not been able to realize until now with sufficiently accurate results. Concurrently, initiatives which focus on the accessibility and regulation of multimedia resources are becoming ever greater in number. Accessibility may be achieved, among other alternatives, by means of the subtitling and audio description of multimedia contents. This paper presents a platform which enables the SA of multimedia content when the subtitling and audio description tasks are being carried out.

KEYWORDS

Semantic annotation, Subtitling, Audio description, Accessible contents, Ontologies

INTRODUCTION

Audiovisual resources have become a significant source of information. Web repositories store millions of hours of videos created by all kinds of users. These platforms represent a new breed of libraries, containing information expressed in formats that differs from the traditional ones but requiring the same level of management involvement. Users accessing a large repository have one common requirement: find what they are looking for. This portrays an important challenge taking into account the vast amount of existing information, which is not always properly organized and structured. The nature of this non-structured information does not facilitate the application of mature techniques such as cataloguing, indexation and recovery, as these frequently rely on text-processing techniques, which are written in natural language. To address this problem, various alternative processing techniques have been proposed. However, such methods have not provided entirely satisfactory results. The aforementioned techniques include manual solutions, such as annotation or categorization of audiovisual resources by means of taxonomies or ontologies; and automatic or semi-automatic solutions, such as image-recognition or sound-processing techniques. In some cases, precision is not the most adequate solution. These solutions are often inefficient as there is a low trade-off between the high cost of the search process from a computational viewpoint, and the output of the results in real time. Additionally, maintaining a multimedia repository implies a costly process. Therefore, providing cataloguing and search tools for audiovisual resources, capable of responding in an efficient and precise manner, arises as a challenge.

On the other hand, accessibility plays a crucial role for people with disabilities. However, IT remains an area lacking consideration towards social access for people. The Web is full of social and cultural opportunities which must be within the reach of all individuals. Web accessibility entails that people with disabilities are able to perceive, understand, navigate, and interact with the Web, and that they can contribute to the Web. In terms of multimedia, the integration of the Internet with other forms of multimedia delivery is just protruding. Audiovisual resources can be considered accessible when they integrate the necessary complements to be "seen" or "heard" by visually or auditory impaired people respectively. Two traditional solutions address these problems: audio description (AD) and closed caption (CC). Both techniques enrich audiovisual resources with new information, and thus furnish new processing possibilities. However, the current rate of adopting these techniques is too low; therefore, the use of the information provided by such techniques is limited and not exploited to its full extent.

This article's approach for cataloguing and processing of audiovisual databases is based on SA of the most common elements that guarantee accessibility to audiovisual resources: CC and AD. Both techniques add two advantages to multimedia resources: the information is textual and therefore, processing it becomes simpler than audiovisual information; the textual information is associated to temporal levels of the audiovisual resource, which makes it possible to work with segments instead of the whole audiovisual resource. As a consequence, this opens new opportunities for carrying out more precise and concrete searches. Association between converting a resource accessible and SA can profit from the context knowledge of captioners and audio descriptors. The person in charge of these tasks should be familiar with the resource and may semantically annotate with little effort subtitles or AD, with scripts being added to the resource in text format. This paper presents SAAMAR, the architecture that supports this proposal.

AUDIOVISUAL ACCESIBILITY

The inclusion of communities of people with disabilities in diverse social and cultural environments is a challenge that should be confronted by society in order to guarantee such groups access to the information. In the domain of multimedia contents, and in particular the video domain, AD and CC represent a medium for those with disabilities to access multimedia environments adequately. Within the multimedia domain, multimedia resources on the Internet should comply with a series of standards. The Web Content Accessibility Guidelines 1.0 [1] is widely considered as a standard by the legislation and regulations of many countries, and its evolution to version 2.0 [2] has demonstrated an advance in diverse aspects of accessibility. Specifically, in the area related to multimedia content, the standard recommends: "Provide a single document that combines text versions of any media equivalents, including captions and AD, in the order in which they occur in the multimedia." It also adds, "Combining text of AD and captions into a single text document creates a transcript of the multimedia, providing access to people who have both visual and hearing disabilities. Transcripts also provide the ability to index and search for information contained in audio and visual materials". A recent and comprehensive overview of multimedia accessibility standards can be found in [3].

Multimedia resources stored on Web sites are generally heterogeneous; however, the majority of them share some characteristics and common problems: they are not accessible by people with disabilities, and processing their information depicts a complex task. The first problem may be solved by means of accessibility tools addition, in the form of CC and/or AD. The second problem concerns the nature of multimedia resources: absence of textual information. Developing a solution to the accessibility problem may in turn provide a solution to the non-textual information processing, as it enriches the resource with such information.

SA

In a domain which is currently increasing in importance, the Semantic Web (SW), the creation of accessible multimedia content which is semantically annotated represents an evolution of the SW concept with regard to the accessibility of its resources. The SA of resources in general, and particularly of multimedia resources, is an arduous task which automatic annotation techniques have not been able to carry out until now with sufficiently accurate results. Concurrently, initiatives focused on the accessibility and regulation of multimedia resources are steadily increasing. Accessibility may be achieved, among other alternatives, by means of the CC and AD of multimedia contents. This paper presents a platform which enables the SA of multimedia content when the CC and AD tasks are being carried out. Through SA, the framework also facilitates the improved retrieval and use of multimedia content by users with disabilities.

CC AND AD

CC consists of a textual transcription of the dialog and contextual sounds that allow people with hearing impairment to read what they cannot hear. CC operation is simple: audiovisual works are divided into temporal segments concerning those fragments with dialog and relevant sounds. These segments are associated with textual transcription which is shown (habitually as a screening over the original image) to users, in such a way that the transcription can be read in the exact moment in which something being captioned can be heard.

The CC process of a work is detailed as follows: temporal segments where subtitles are required are defined, and the corresponding transcription or contextual information is integrated into each one of them. There are tasks that use CC in order to carry out SA. Using CC has disadvantages. The most important is that most of the information is about dialog [4] and does not capture much of the information that is being presented in videos. More information about what is being watched is required. AD could represent a solution for this problem.

AD provides a similar function for visually impaired people. In this case, temporal segments should fit the "white" gaps, in other words, those fragments without dialog. Taking advantage of these gaps, a voice-over completes the dialog information by means of a locution that transforms visual information into auditory information.

The AD process of an audiovisual resource is similar to that of a CC, with format differences. In this case, first of all, temporal segments where AD may be included are defined. Hereafter, a script should be prepared. This script should sustain the locution to be integrated to the soundtrack, matching the established temporal segments.

Both processes are quite complex since they should contemplate certain technical and formal rules in order for the result to be useful. Nonetheless, in both cases there is an important coincidence: the association of textual information (transcription in the case of CC or script in the case of AD) to temporal segments in the work (whose ranges are expressed according to EBU format, in the form HOUR:MINUTE:SECOND:FRAME) as shown in Figure 1.

Figure 1

Figure 1. CC and AD diagram

ACCESSIBILITY AND SA: BRIDGING THE GAP

Our goal is to harness the processes of subtitling and AD of audiovisual resources to perform a semantic labeling of these resources. This approach establishes a relationship between time segments and semantic information, rather than label all audiovisual resources, using the complete resource as the smallest unit of information to process semantically. It also aims to exploit the contextual knowledge of users who subtitle and AD of resources as an alternative to carrying out this activity in two steps: subtitles (or AD) and labeling.

The first step is to determine the time segments. The segments should not overlap in any case, whether these are created to accommodate captions or AD. For each time segment, it will include the text of subtitling or AD script accordingly. In our proposal, the process will be linked to a contextual ontology, and during the insertion of text, the system will propose a SA to some entries, correspondence with classes or entities of the ontology. Alternatively, the user may decide if a word must be semantically tagged and can do so without the system having made the proposal earlier.

The result of this process shall be as shown in the example below:

Start segment	End segment	CC or AD	Caption line or Audio description script	Semantic labeling
00:05:02:20	00:05:10:15	CC	*I like this building. It brings back good memories.*	
00:05:30:01	00:05:37:10	AD	*The protagonist stands with his back to Notre Dame*	<French_Gothic>Notre Dame</French_Gothic>
00:05:48:18	00:05:51:14	CC	*Why?*	
00:06:01:01	00:06:01:12	CC	*Can you see that tower? It's called the tower of Saint-Romain.*	<Early_Gothic>tower of Saint-Romain</Early_Gothic>

Table 1. Semantic Annotation

In regard to the textual information included as a consequence of adding subtitles or AD scripts, a SA is performed. It profits from the conversion process of an accessible audiovisual resource and the contextual knowledge of users in charge of the process. The result of the given example is the link between concrete temporal segments and architectonic characteristics as shown in Figure 2.

Figure 2

Figure 2. Semantic labelling activity

THE SAAMAR APPROACH

Web sites for video sharing store hundreds of thousands of small-scale amateur works, film trailers, music videos, or commercials of only few seconds in length, and most of them are not accessible. There are diverse reasons for this lack of accessibility, ranging from ignorance to profitability issues. Search engines which index these databases are based on metadata, and the results are not always satisfactory. Multimedia resources may be accessible if they provide alternative methods of access to their contents for people with disabilities. In order to achieve accessibility, various solutions and tools exist, corresponding to the type of resource, to provide accessibility to the content. Thus, for audio resource we can use caption or transcription and for image resources, alternative text. This paper is focused on video resources, using CC and AD as accessibility tools.

SA during a transformation process of audiovisual resources to accessible resources implies the integration of two required activities of a different nature. Our proposal consists of defining an architecture that fulfils every requirement in order to carry out both activities jointly.

Systems based on this architecture should enable CC and AD of audiovisual resources. The main problem found with these two activities is to establish temporal segments adequately. Users should indicate the beginning and end of the segment, and the system should prevent overlaps by means of timing chain control.

Several possible solutions have been found to solve the requirement of a semantic information base. Here, the utilization of contextually enclosed ontologies is proposed. However, some other semantic structures, such as thesauri and taxonomies, may be used. In any case, this architecture should contemplate a semantic assistant. This assistant should narrow the gap between the user interface used to insert text (subtitles or AD scripts) and the semantic information base. This semantic assistant may confirm whether a term inserted in the system as a part of CC or AD, is present in the referenced ontology. It may also propose alternative SAs. This process is semiautomatic; the system proposes annotation alternatives but is the user who determines whether annotation is required, and resolves possible ambiguities. This implies that, for an accurate annotation, users (transcribers) should improve both their SA and captioning skills. This circumstance could, however, result in an increase in costs, and due to this in a barrier to the adoption of the tool and must be solved in future work.

The result of semantically-annotated CC and AD is a relation of temporal segments which “talk about” or “present” concrete terms within the selected ontology (or chosen semantic medium). It establishes a synchronized relationship between temporal segments and “objects”, similar to that defined by SMIL [5] and Daisy [6]. In practice, annotation enables querying the system through the implementation of this kind of architecture, with requests such as “list of sequences in which a gothic cathedral appears in this video” or “the dialog fragments that mention ancient gothic architectonic elements”. Possibilities are as varied as the complexity of the SA executed.

Our proposal consists of the definition of a semantic-annotation oriented architecture of short videos. This annotation is carried out during CC and AD processes of videos, providing semantic support based on contextually enclosed ontologies.

THE SAAMAR ARCHITECTURE

SAAMAR was developed as a result of a series of specific requirements. Other implementations currently available partly fulfill these requirements. However, the motivation for SAAMAR was to cover all requirements in a single platform.

- Accessible multimedia resources. The architecture should provide the necessary tools to grant accessibility to resources, and these resources should be designed to fulfill a number of other remaining requirements.
- Chronological boundaries. The SA of multimedia elements has some variations based on the element type. A still image contains a textual description linked to the image. In an image containing movement or a video, CC, AD or transcriptions are linked to a time sequence, given that each “frame” or fragment of audio or video is always accompanied by its associated informative text. In the case of providing accessibility to a multimedia element which is not a still image, this element should be divided into fragments. In the case of transcription or CC, a fragment of sound with a specific temporal segmentation has an associated text which is displayed to the user during this particular time segment. Similarly, in the case of AD, the segments of the multimedia elements termed “empty” (gap spaces without conversation or relevant sounds) have an associated sound element which explains the occurrence of the sound gap in order to complete the information perceived by people with visual disabilities. This sound element is an additional multimedia element which presents annotation difficulties that the current work proposes to resolve. Fortunately, in the current context, the solution is obvious, given that this sound element is a section of the previously written AD script, which can be linked to the element and can thus be annotated. Therefore, except in the case of still images, the annotator should have a series of elements available for the management of temporal segments, which are associated with these temporal segments.
- Semantic Support. The semantic base is a set of categorized and related elements which support the annotation process. We use an ontology written in OWL language to support the creation of semantic tags, using OWL language, but this metadata could be expressed with another semantic structure. Selection of metadata annotation from a support element comprised of a semantic base avoids these types of errors, and provides the guarantee that the metadata creation is correct. Without a semantic base, a question arises: How can context be retrieved based on indirect queries?
- SA based on contextual support. SAAMAR should offer SA support within the context of the resource which is being made accessible. It should include the functionality of automatic access to the semantic support in real time, to provide annotation

alternatives based on the content currently being listened to or viewed. The user (annotator) should receive proposals for annotations based on the semantic support used. This ensures that the annotation is carried out alongside the creation of the accessibility elements, and that this process is efficient and precise.

- Retrieval based on semantic technology (ST). ST enabling interoperability represents a significant improvement in data search and retrieval. Thus, retrieval techniques based on ST offer added value, determining relationships during searches which are not included in metadata, oriented towards the specification of elements which form multimedia contents at any granularity level.

SAAMAR, as depicted in Figure 3, is divided into high-level subsystems, which are comprised of the following:

1. Multimedia Metadata Description Standard. It is a software component itself with regard to the support data and infrastructure, and the software access mechanisms.
2. Audio description/Caption/Transcription interface. This interface will be employed by users to insert the accessibility mechanisms (AD, Caption, Transcription) into the metadata. By means of this subsystem, the "population" of the metadata is carried out, aided by the Semantic Engine and Semantic Data Support components. The Semantic Engine module provides module options for labelling semantic information related to terms which have been entered to the interface. The user selects the most appropriate option and stores the annotation using the Multimedia Description Standard Metadata.
3. Semantic Engine. This module uses ST in both of the operating constituents in which it is applied. On the one hand, it provides access to the Semantic Support module, with the aim of supplying classes or instances contained in the module with objects for formal metadata development. In the other context, dedicated to access, it provides the necessary help to carry out a semantic search over the metadata, without limiting itself to the traditional "literal" search.
4. Semantic Data Support. This module contains the ontology, taxonomy, or any other tool for semantic representation of knowledge. This module is only structural, in contrast with the Multimedia Metadata Description Standard, which includes the data access software components in its definition. In the case of Semantic Data Support, this data access is located within the Semantic Engine.

Figure 3

Figure 3. SAAMAR architecture

EVALUATION

Experiments

We have developed a Java based prototype to validate the architecture. Using this prototype, a user can load a video, select the segments over which the CC will be visualized, and insert them. SAAMAR assists the transcriber by proposing the annotation alternatives (the tokens) which have correspondence in the ontology contained in the Annotation Support component, in terms of context. The user should only select the alternative which corresponds to the meaning of the token introduced. This semi automatic annotation mechanism brings the opportunity to annotate CC taking full advantage of ST, but also to perform this time consuming task in a controlled and assisted way. Using SAAMAR, the user will indicate the time segments into which the embedded CC will be inserted. SAAMAR enables simple SA, from a list of proposed concepts taken from the ontology. Users are provided with semantic information to add; therefore, while editing a caption, they are able to annotate a word or a set of words with semantic data, just as easy as marking the selected words and associating them with a property or vocabulary concept from the ontology domain. For the purpose of this work the mechanism chosen for annotating is sufficient and satisfactory.

With the objective of carrying out an empirical evaluation of the results of the platform use, testing of the platform was performed in a defined environment. A multimedia format was utilized which constituted audio and video of fifty-five seconds duration, a similar length to television commercials. The experiment consisted of carrying out CC, AD and the later SA of content in two distinct scenarios. In the first place, the researchers performed AD and CC of the multimedia content using the AEGISUB tool, and later users were asked to perform the SA of the contents manually. Secondly, SAAMAR was used to carry out the same task.

With the objective of comparing the results of the evaluation with a standard, a group of experts agreed upon a SA upon which consensus was achieved among all the experts. This annotation was established by a set of experts using the DELPHI method based on the viewing of the multimedia format, individually in the first place, in order to achieve group consensus subsequently.

The experimentation had a double objective. The first objective was to determine if SAAMAR provides increased utility to the user with regard to carrying out the joint tasks of captioning and AD. The first objective of the evaluation was achieved by administering a questionnaire to the subjects who carried out the experiment. In order to complete the evaluation, the subjects were requested to indicate their level of agreement with the following statements. 1) SAAMAR is a useful tool for completing the tasks required. 2) SAAMAR is a tool which speeds up the work. 3) SAAMAR is a tool which adds convenience to the process. The responses to the questions were codified by the users on a Likert scale ranging from 1-5, with the following values: 1. Strongly disagree; 2. Disagree; 3. Neutral; 4. Agree; 5. Strongly agree.

The second objective was to establish whether the results of SA are more satisfactory as a result of using SAAMAR than the results obtained from another technique. To perform this test, the results of the SA of both scenes were compared with the standard annotation obtained.

Sample

The sample comprised 12 individuals skilled in CC and AD tasks, with seven women and five men. The average age of the subjects was 27.8. All of the subjects had similar experience with captioning technologies; however, they did not have experience in SA of digital contents. The tasks were performed individually by each subject, who was isolated from the rest of the group during the completion of the tasks. All of the annotation tasks were carried out during December 2008.

Additionally, the sample subjects who applied the DELPHI method in order to establish the SA to be considered as standard comprised three males, with an average age of 32.3. All of them can be considered as experts in semantics and SA.

Results

The results in relation to the acceptance of SAAMAR are highly satisfactory. The application of the questionnaires to the subjects has produced the results which are displayed in Figure 4:

Figure 4

Figure 4. Questionnaire results

All of the opinions in relation to SAAMAR are positive, with different levels of agreement among the subjects being shown. Examining the results, it is evident that SAAMAR is a valid alternative for the double task of annotating and captioning and AD. In particular, it is especially notable the fact that 83% of users consider SAAMAR as a much faster valid alternative, and 75% of users consider the tool convenient for the process. None of the evaluations of SAAMAR resulted negative in terms of the aspects which made up the questionnaire.

It was also considered interesting to analyze the results of the annotation process carried out. To perform the analysis, the annotation carried out by the experts using the DELPHI method was selected as a base, and compared with the annotations chosen by the users. The experts, using DELPHI method, defined a total of 13 correct SAs for the multimedia clip. These SA defined by experts will be used as the correct pattern for the evaluation of SAAMAR. In tests with SAAMAR, the users generated a total of 126 annotations, and using the integrated captioning option produced a total of 104. To verify that annotations were correct, researchers decided that each annotation should semantically represent the object required, and should do so at the correct instant, establishing a margin of +-2 seconds for acceptance. Applying these parameters, of the 126 SAAMAR annotations, 117 were correct, meanwhile of the 104 annotations using the other method, only 83 were correct. The 9 errors produced by the SAAMAR annotations were due to incorrect semantic identification, while using the other method, 10 errors were semantic and 11 were related to timing.

The first conclusion about this experimentation is the increase of annotations produced by SAAMAR users. Thus, SAAMAR users produced 10.5 annotations per subject, while following the DELPHI method, users produced just 8.7 annotations per subject. This variability among participants is grounded in the integrated nature of SAAMAR. Given that SAAMAR implements an architecture in which CC and SA are performed at the same time by suggesting SAs, these results confirm that this approach brings out higher annotation density marks. A preliminary analysis of the data reveals also that the annotations carried out using SAAMAR are more accurate. However, a more comprehensive analysis was also considered necessary. To evaluate the performance of annotation of both environments, the standard recall, precision and F1 measures were applied. Recall and Precision measures reflect the different aspects of annotation performance. The F1 measure was later introduced in order to combine precision and recall measures, with equal importance, into a single parameter for optimization. All results of SAAMAR (Precision=0.93, Recall = 0.75, F1 = 0.83) are higher than the other annotation technique (Precision = 0.80, Recall = 0.53, F1 = 0.64).

A brief analysis of the metrics confirms the utility of SAAMAR. Recall displays a slightly lower value, which may be due to the fact that the annotation taken as a base was very exhaustive. In all cases, the evaluation results of SAAMAR are better than those achieved by combined annotation. This circumstance verifies the synergy which SAAMAR aims to exploit. Thus, the description provided in the manual captioning and AD is enriched both in terms of the quantity of SAs as well as their quality. This process assumes an increase in the value of annotation, transforming multimedia contents into elements which are more easily referenced and thus accessible.

RELATED WORK

The benefit of adding semantics to any content consists of bridging nomenclature and terminological inconsistencies to include underlying meanings in a unified manner. In order to achieve the concept described by "*semantic content*", it is necessary for resources to be associated with metadata. Since metadata generated by automated support tools is error-prone and often requires correction [7], a safer mechanism for associating such metadata is annotation. According to the New Oxford Dictionary of English, annotation is "a note by way of explanation or comment added to a text or diagram". SA goes beyond familiar textual annotations about the content of documents; it formally identifies concepts and relationships between concepts in documents, and is intended primarily for use by machines [8]. Unlike an annotation in the normal sense, a SA must be explicit, formal, and unambiguous: explicit makes a SA publicly accessible, formal makes a SA publicly agreeable and unambiguous makes a SA publicly identifiable [9].

SA has two additional benefits when compared to metadata annotation: enhanced information retrieval and improved interoperability [8]. In spite of the advantages of SA, a potential barrier to the uptake of ST is the effort required to mark up information with SAs [10]. Annotation tools may be categorized into several types: manual, semi-automatic or automatic. As mentioned above, automatic annotation tools present inaccuracies with regard to error occurrences [7], while manual annotation similarly presents a drawback, but in the sense of it being a costly, time-consuming process.

High-quality metadata is essential for multimedia applications [11]. Taking into account that the high quality of annotations can be guaranteed with the ontologies used, there are many works which discuss the use of multimedia SA based on ontologies. In the SA of multimedia content fields, there are some works [12], [13] designed to add semantics to multimedia content by using tools and ontologies. However, none of the previous efforts have focused on the use of SA of multimedia content combined with the consideration of the accessibility of the content. SAAMAR proposes using the AD and captioning processes to carry out semi-automatic SA. In this way, it is aimed to minimize the problems of the inefficient speed of SA, by applying the process during AD and captioning. This process has the advantage of greater speed and precision of the final process, which results in improved search and retrieval of information.

CONCLUSIONS

This paper has presented the SAAMAR architecture, developed as a tool for SA of accessible multimedia resources, basing itself on a standard for multimedia information. The objective of the work was to demonstrate the viability of carrying out assisted SA using contextual help systems during the captioning, AD or transcription of multimedia resources, which has been achieved. Additionally, as part of the developed proposal, an evaluation of the results has been completed from a formal perspective. As a result of the evaluation, satisfactory indicators have been obtained; SAAMAR is faster and obtains better annotation results when compared with the alternative of using independent annotation tools. This improvement in results is evident in relation to the reduction of errors in the SA of multimedia resources. The results are accompanied by a high level of acceptance of the tool subsequent to testing by a group of users.

As future work, four different research lines are proposed. Firstly the authors propose to design an annotation interface which exploits new ways for video annotation, and additionally, capacities of the semantic search such as faceted search. This new redesign should also be focused on the user interface for a smoother adoption of the tool by, among other factors, cutting training costs. Secondly, it is proposed to extend the current experimentation and the tool itself in order to span multimedia contents of longer duration. This change in focus will allow the tool to become an alternative for conventional annotation, accessing a much larger vocabulary for this task, which would present a challenge for the selection of applicable ontologies. Thirdly, given the limited study samples presented in this paper, the authors propose a wider experimental setup in order to include qualitative experimentation. Regarding the limited amount of audio transcribers and accessible multimedia analyzed, the authors believe that this new and complementary approach may also portray a contribution to the existing literature as well as an increase in the system applicability. Finally, in a purely experimental scenario, it is proposed to measure the time required for annotation, to further study the tool performance.

REFERENCES

1 W3C, Web Content Accessibility Guidelines 1.0 (WCAG 1.0) (1999), http://www.w3.org/WAI/intro/wcag.php

2 W3C, Web Content Accessibility Guidelines 2.0 (WCAG 2.0) (2008), http://www.w3.org/WAI/intro/wcag.php

3 Moreno, L., Martínez, P. & Ruiz-Mezcua, B. (2008). Disability Standards for Multimedia on the Web. IEEE Multimedia, 15(4), 52-54.

4 Brezeale, D. & Cook, D. J. (2009) Learning Video Preferences Using Visual Features and Closed Caption. IEEE Multimedia. 16(3), 39-47.

5 Bulterman, D., et al. (2008). Synchronized Multimedia Integration Language (SMIL 3.0) Specification. http://www.w3.org/TR/SMIL3/ .

6 DAISY Consortium (1996). Digital Audio-based Information System, http://www.daisy.org.

7 Gennaro, C. (2008). Regia: a metadata editor for audiovisual documents. Multimedia Tools and Applications, 36 (3), 185-201.

8 Uren, V.S., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E. & Ciravegna, F. (2006). Semantic annotation for knowledge management: Requirements and a survey of the state of the art. Journal of Web Semantics, 4(1), 14-28.

9 Ding, Y., Embley, D.W. & Liddle, S.W. (2006). Automatic Creation and Simplified Querying of Semantic Web Content: An Approach Based on Information-Extraction Ontologies. In Mizoguchi, R., Shi,Z. & Giunchiglia, F (Eds.), ASWC 2006, LNCS 4185, (pp. 400–414). Berlin / Heidelberg: Springer.

10 Benjamins, V.R., Davies, J., Baeza-Yates, R., Mika, P., Zaragoza, H., Greaves, Gómez-Pérez, J.M., Contreras, J., Domingue, J. & Fensel, D. (2008). Near-Term Prospects for Semantic Technologies. IEEE Intelligent Systems, 23(1), 76-88.

11 Nack, F., van Ossenbruggen, J. & Hardman, L. (2005). That Obscure Object of Desire: Multimedia Metadata on the Web (Part II). IEEE Multimedia, 12(1), 54-63.

12 Nagao, K., Shirai, Y. & Squire, K. (2001). Semantic annotation and transcoding: making Web content more accessible. IEEE MultiMedia, 8(2), 69-81.

13 Park, K.W., Jeong, J.W. & Lee, D.H. (2007). OLYBIA: Ontology-Based Automatic Image Annotation System Using Semantic Inference Rules. In: Proceedings 12th International Conference on Database Systems for Advanced Applications, Bangkok, Thailand.

Authors Bios

Fernando Paniagua Martín is a Teaching Assistant at the Carlos III Technical University of Madrid. Currently, he is completing his PhD in Computer Science in the Universidad Carlos III de Madrid. He also holds a Master in Computer Science and Technology. His research interests include Audio-visual accessibility on the Web and Web 2.0 & 3.0.

Angel García-Crespo is the Head of the SofLab Group at the Computer Science Department in the Universidad Carlos III de Madrid and the Head of the Institute for promotion of Innovation

Pedro Juan de Lastanosa. He holds a PhD in Industrial Engineering from the Universidad Politécnica de Madrid and received an Executive MBA from the Instituto de Empresa. He is the author of more than a hundred publications.

Ricardo Colomo-Palacios is an Associate Professor at the Computer Science Department of the Universidad Carlos III de Madrid. His research interests include applied research in Information Systems, Software Project Management and Social and Semantic Web. He received his PhD in Computer Science from the Universidad Politécnica of Madrid (2005). He also holds a MBA from the Instituto de Empresa (2002).

Belen Ruiz-Mezcua is a Lecturer at the Computer Science Department of the Universidad Carlos III de Madrid. She holds a PhD in Sciences Physics from the Telecommunications School of Polytechnic University of Madrid. She is Vice-chancellor of Research adjunct to the Scientific Park. Her research interests are focused on Biometrics and speech processing and artificial intelligence.

Contact Information

Fernando Paniagua-Martín, fernando.paniagua@uc3m.es, +34 91 624 5962

Ángel García-Crespo, angel.garcia@uc3m.es, +34 91 624 9417

Ricardo Colomo-Palacios*, ricardo.colomo@uc3m.es, +34 91 624 5958

Belén Ruiz-Mezcua, mbelen.ruiz@uc3m.es, +34 91 624 9968

Computer Science Department,

Universidad Carlos III de Madrid

Av. Universidad, 30.

28911 Leganés (Madrid). Spain

Fax: +34 91 624 9129