Semantic Annotation Architecture for Accessible Multimedia Resources

Fernando Paniagua-Martín, Ángel García-Crespo, Ricardo Colomo-Palacios, and Belén Ruiz-Mezcua
Universidad Carlos III de Madrid

A platform enables the semantic annotation of multimedia content when the subtitling and audio description tasks are being carried out.

Audiovisual resources have become a significant source of information. Web repositories store millions of hours of videos created by all types of users. These platforms represent a new breed of libraries, containing information expressed in formats that differ from traditional formats but requiring the same level of management involvement. Users accessing a large repository have one common requirement, to find what they are looking for, which represents an important challenge that must take into account the vast amount of existing information that is not always properly organized and structured.

The nature of this nonstructured information doesn't facilitate the application of mature techniques, such as cataloging, indexing, and recovery, as these frequently rely on text-processing techniques that 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 semiautomatic solutions, such as image-recognition or sound-processing techniques. These solutions are often inefficient as there is a 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 is a costly process, and providing audiovisual cataloging and search tools, capable of responding efficiently and precisely, is a challenge.

In addition, accessibility plays a crucial role for people with disabilities. Accessibility means that people with disabilities are able to perceive, understand, navigate, and interact with the Web. In terms of multimedia, the integration of the Internet with other forms of multimedia delivery is just beginning. Audiovisual resources can be considered accessible when they integrate the necessary complements to be seen or heard by visually- or hearing-impaired people. Two traditional solutions address these problems: audio description (AD) and closed captioning (CC). Both techniques enrich audiovisual resources with new information. 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 cataloging and processing audiovisual databases is based on semantic annotation (SA) of the most common elements that guarantee accessibility to audiovisual resources through CC and AD. Both techniques add advantages to multimedia resources: the information is textual and therefore processing it becomes simpler than processing audiovisual information; the textual information is associated with temporal aspects 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. To address these considerations, we have developed the semantic annotation accessible multimedia resources (Saamar) architecture.

Audiovisual accessibility

The inclusion of communities of people with disabilities in diverse social and cultural environments is a challenge that should be confronted by society to guarantee such groups access to the information. In the domain of multimedia content, and in particular the video domain, AD and CC represent a medium for those with disabilities to access multimedia environments. 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 providing ''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 recommends combining text of AD and captions into a single text document to create a transcript of the multimedia to provide access to people who have both visual and hearing disabilities. The standard notes that transcripts 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 elsewhere.[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 is complex. The first problem can be solved by means of accessibility tools in the form of CC and AD. The second problem concerns the nature of multimedia resources: absence of textual information. Developing a solution to the accessibility problem might in turn provide a solution to the nontextual information-processing issue.

In a domain that is currently increasing in importance, the Semantic Web, the creation of accessible multimedia content that is semantically annotated represents an evolution of accessibility. The SA of resources in general, and of multimedia resources in particular, is an arduous task that automatic annotation techniques haven't 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 might be achieved, among other methods, by means of the CC and AD of multimedia content. The Saamar platform 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.

Closed captioning and audio description

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

While there are tasks that use CC to carry out SA, using CC has disadvantages. The most important is that most of the information is about dialog[4] and doesn't capture much of the information being presented in videos. More information about what is being watched is required. AD could represent a solution to this problem. AD provides a similar function for visually-impaired people. In this case, temporal segments fit the white gaps, 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. First of all, temporal segments where AD might be included are defined. Hereafter, a script sustains the locution to be integrated to the soundtrack, matching the established temporal segments.

Both processes are quite complex because they must deal with certain technical and formal rules 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 the subtitling data exchange format from the European Broadcasting Union (EBU), in the form hour:minute:second:frame) as shown in Figure 1 (next page).

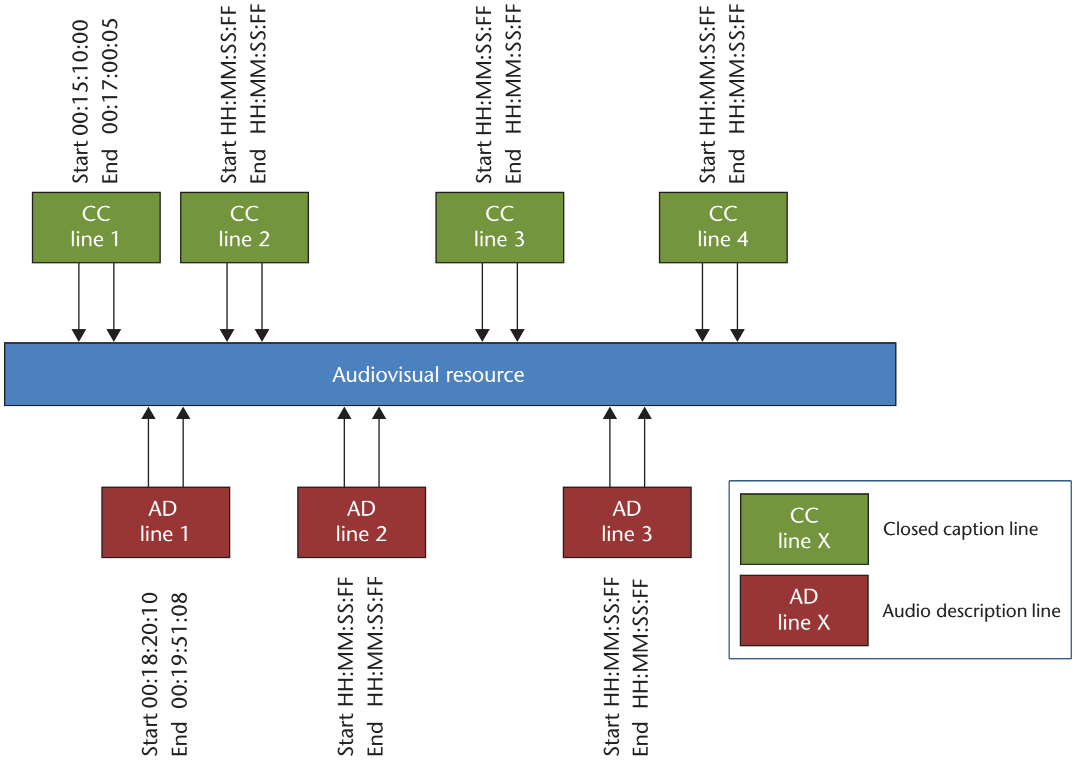

Figure 1. Closed captioning and audio description diagram.

Table 1. Semantic annotation.

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

Accessibility and semantic annotation

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. 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 is linked to a contextual ontology, and during the insertion of text, the system proposes an SA to some entries, corresponding to classes or entities of the ontology. Alternatively, the user in charge of the annotation process might decide if a word must be semantically tagged and can do so without the system having made the proposal earlier. Table 1 shows the result of this process.

In regard to the textual information included as a consequence of adding subtitles or AD scripts, an SA is performed, profiting from the conversion process of an accessible audiovisual resource and the contextual knowledge of the 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.

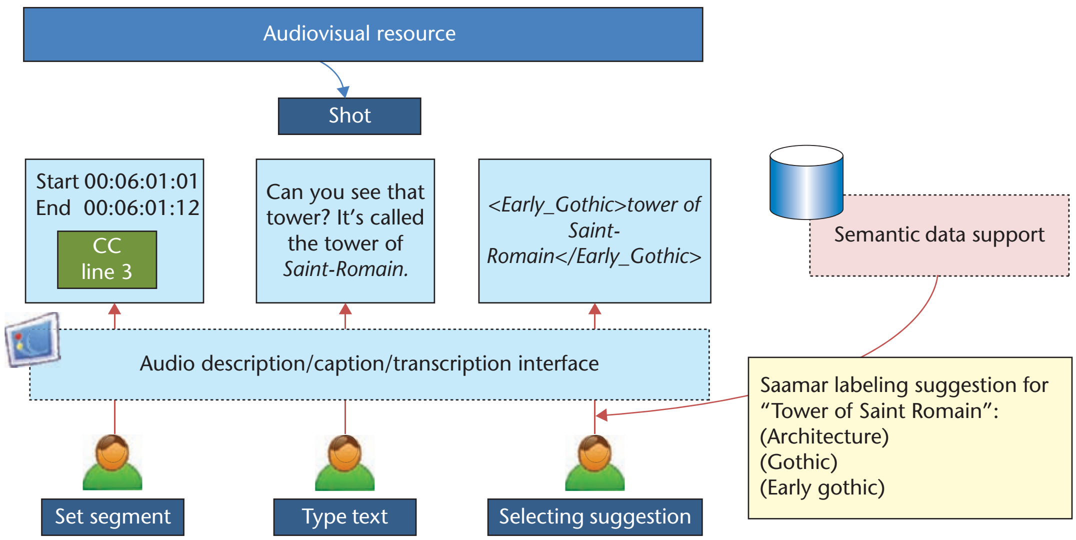

Figure 2. Semantic labelling activity.

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 a few seconds in length. Most of these works are not designed for accessibility. There are diverse reasons for this lack of accessibility, ranging from ignorance to profitability. Search engines that index these databases are based on metadata, and the results are not always satisfactory. Multimedia resources might be accessible if they provided alternative methods of access to their content for people with disabilities. For audio resources, we can use captions or transcription, and for image resources, alternative text. This article focuses on video resources, using CC and AD as accessibility tools.

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

The result of semantically annotated CC and AD is a relation of temporal segments that talk about or present concrete terms within the selected ontology (or chosen semantic medium). This method establishes a synchronized relationship between temporal segments and objects, similar to that defined by Synchronized Multimedia Integration Language[5] and Daisy (see http://www.daisy.org). 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 according to a series of specific requirements. While other implementations partly fulfill these requirements, the motivation for Saamar was to cover all requirements in a single platform. The architecture provides the necessary tools

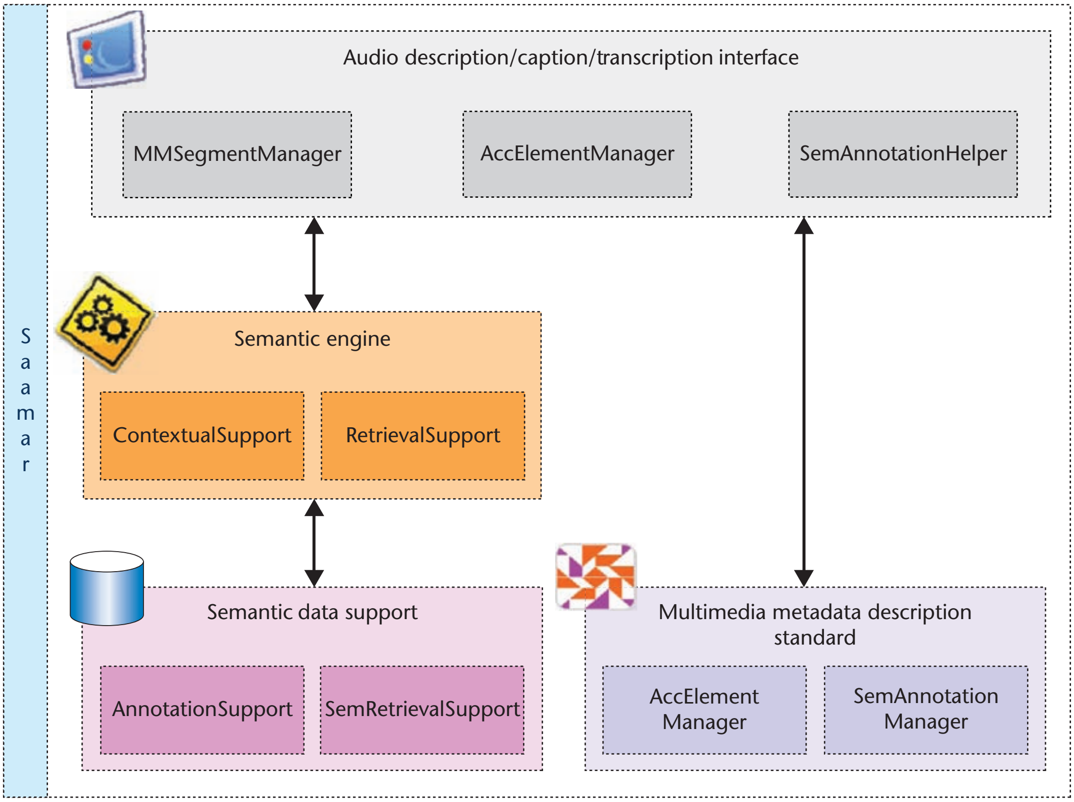

Figure 3. Saamar architecture.

to grant accessibility to resources; these resources should be designed to fulfill a number of other remaining requirements.

The SA of multimedia elements varies by 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 clip is always accompanied by its associated informative text. For providing accessibility to a multimedia element that isn't a still image, the element must be divided into fragments. In the case of transcription or CC, a fragment of sound with a specific temporal segmentation has an associated text that 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 that explains the occurrence of the sound gap to complete the information perceived by people with visual disabilities.

This sound element is an additional multimedia element that presents annotation difficulties that our proposed system resolves. 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 must have a series of elements available for managing of temporal segments.

The semantic base is a set of categorized and related elements that support the annotation process. We use an ontology written in Web Ontology Language (OWL) to support the creation of semantic tags Selection of metadata annotation from a support element comprised of a semantic base helps avoid errors Saamar offers SA support within the context of the resource being made accessible. It includes automatic access to the semantic support in real time, to provide annotation alternatives for the content currently being listened to or viewed. The user (annotator) receives proposals for annotations on the basis of the semantic support used, ensuring that the annotation is carried out, efficiently and precisely, alongside the creation of the accessibility elements.

Retrieval techniques based on semantic technology offer added value, determining relationships that are not included in metadata. Saamar, as depicted in Figure 3, is divided into high-level subsystems, which are comprised of the following:

- *Multimedia metadata description standard.* A software component that supports the data, infrastructure, and software-access mechanisms.

- *Audio description/caption/transcription interface.* An interface for inserting the accessibility mechanisms 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 provides module options for labelling semantic information related to terms entered into the interface. The user selects the most appropriate option and stores the annotation using the multimedia metadata description standard.

- *Semantic engine.* A module that 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. It also provides the necessary help to carry out a semantic search without being limited to traditional literal search.

- *Semantic data support.* A module that 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.

Evaluation

We developed a Java-based prototype to validate the architecture. With this prototype, a user can load a video, select the segments over which the captions will be visualized, and insert them. Saamar assists the transcriber by proposing the annotation alternatives (the tokens) that contextually correspond to the ontology contained in the annotation support component. The user selects the alternative that corresponds to the meaning of the token introduced. This semiautomatic annotation mechanism allows for annotating CC to take advantage of semantic technology, but also to perform this time-consuming task in a controlled and assisted way.

With Saamar, the user indicates the time segments into which the embedded closed caption 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 easily 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 platform use, testing of the platform was performed in a defined environment. We used a multimedia format that constituted audio and video lasting 55 seconds, 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 part, we performed AD and CC of the multimedia content using the Aegisub subtitle tool (see http://www.aegisub.org/), and then users were asked to perform the SA of the content manually. In the second part, 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 on an SA. This annotation was established by experts using the Delphi method based on viewing the multimedia format, individually in the first place, to achieve group consensus.

The experiment had a double objective. The first objective was to determine whether Saamar provides increased utility to the user with regard to carrying out the joint tasks of captioning and AD. The first objective was achieved by administering a questionnaire to the subjects who carried out the experiment. To complete the evaluation, the subjects were requested to indicate their level of agreement with the following statements:

- Saamar is a useful tool for completing the tasks required.

- Saamar is a tool that speeds up the work.

- Saamar is a tool that adds convenience to the process.

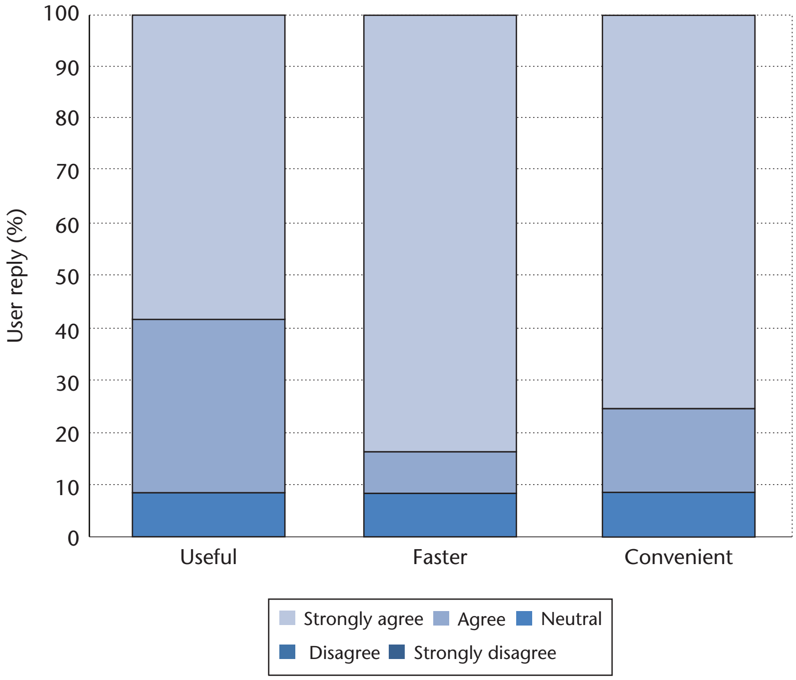

Figure 4. Questionnaire results.

The responses to the questions were codified by the users on a Likert scale ranging from 1 to 5, with the following values: 1 strongly disagree; 2 disagree; 3 neutral; 4 agree; and 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 didn't have experience in SA of digital content. 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 to establish the SA to be considered as standard comprised three males, with an average age of 32.3. All of them could be considered experts in semantics and SA.

Results

The results in relation to the acceptance of Saamar were highly satisfactory. The application of the questionnaires to the subjects produced the results shown in Figure 4.

All of the opinions in relation to Saamar were positive, with different levels of agreement among the subjects. Examining the results, it's evident that Saamar is a valid alternative for the double task of annotating and captioning and AD. In particular, it's especially notable that 83 percent of users considered Saamar to be a much faster valid alternative, and 75 percent considered the tool to be convenient for the process. None of the evaluations of Saamar had negative results in terms of the aspects that made up the questionnaire.

In addition, we analyzed the results of the annotation process. 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 the Delphi method, defined a total of 13 correct SAs for the multimedia clip. These SAs defined by experts were 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 nine 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.

This experiment showed an increase of annotations produced by Saamar users. 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

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. To achieve the concept described by *semantic content*, it's necessary for resources to be associated with metadata. Because metadata generated by automated support tools is error-prone and often requires correction,[1] 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.'' Semantic annotation (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.[2] Unlike an annotation in the normal sense, an SA must be explicit, formal, and unambiguous: explicit makes an SA publicly accessible, formal makes an SA publicly agreeable, and unambiguous makes an SA publicly identifiable.[3]

SA has two additional benefits when compared to metadata annotation: enhanced information retrieval and improved interoperability.[2] In spite of the advantages of SA, a potential barrier to the uptake of semantic technology is the effort required to mark up information with SAs.[4] Annotation tools might be categorized into several types: manual, semiautomatic, or automatic. Automatic annotation tools present inaccuracies with regard to error occurrences,[1] 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.[5] Taking into account that the high quality of annotations can be guaranteed with the ontologies used, there are many works that discuss the use of multimedia SA based on ontologies. In the SA of multimedia content fields, there are some works[6,7] 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. The semantic annotation architecture for accessible multimedia resources (Saamar) is designed to use the audio description and captioning processes to carry out semiautomatic SA. In this way, the approach aims to minimize the problems of the inefficient speed of SA by applying the process during audio description 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.

References

1. C. Gennaro, ''Regia: A Metadata Editor for Audiovisual Documents,'' *Multimedia Tools and Applications,* vol. 36, no. 3, 2008, pp. 185-201.
2. V.S. Uren et al., ''Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art,'' *J. Web Semantics,* vol. 4, no. 1, 2006, pp. 14-28.
3. Y. Ding, D.W. Embley, and S.W. Liddle, ''Automatic Creation and Simplified Querying of Semantic Web Content: An Approach Based on Information-Extraction Ontologies,'' *Proc. 1st Asian Semantic Web Conf.,* R. Mizoguchi, Z. Shi, and F. Giunchiglia, eds., LNCS 4185, Springer, 2006, pp. 400-414.
4. V.R. Benjamins et al., ''Near-Term Prospects for Semantic Technologies,'' *IEEE Intelligent Systems,* vol. 23, no. 1, 2008, pp. 76-88.
5. F. Nack, J. van Ossenbruggen, and L. Hardman, ''That Obscure Object of Desire: Multimedia Metadata on the Web (Part II),'' *IEEE MultiMedia,* vol. 12, no. 1, 2005, pp. 54-63.
6. K. Nagao, Y. Shirai, and K. Squire, ''Semantic Annotation and Transcoding: Making Web Content More Accessible,'' *IEEE MultiMedia,* vol. 8, no. 2, 2001, pp. 69-81.
7. K.W. Park, J.W. Jeong, and D.H. Lee, ''Olybia: Ontology-Based Automatic Image Annotation System Using Semantic Inference Rules,'' *Advances in Databases: Concepts, Systems and Applications,* vol. 4443, 2007, pp. 485-496.

marks. A preliminary analysis of the data reveals 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 reflected the different aspects of annotation performance. The F1 measure was later introduced 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) were higher than the other annotation technique (precision = 0.80, recall = 0.53, F1 = 0.64).

A brief analysis of the metrics confirmed the utility of Saamar. Recall displays a slightly lower value, which may have been due to the fact that the annotation taken as a base was exhaustive. In all cases, the evaluation results of Saamar were better than those achieved by combined annotation. These findings verify the synergy that Saamar is designed 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 content into elements that are more easily referenced and thus accessible.

Conclusions

For our future work, we are planning four different research lines. First, we aim to design an annotation interface that exploits new methods for video annotation and new capacities of semantic search, such as faceted search. This new redesign, which will focus on the user interface for a smoother adoption of the tool, should help cut training overhead. Second, we plan to extend the current experimentation and the tool itself to span multimedia content of longer duration. This change in focus will allow the tool to become an alternative for conventional annotation by accessing a much larger vocabulary that would present a challenge for the selection of applicable ontologies. Third, given the limited study samples presented in this article, we plan a wider experimental setup to include qualitative experimentation. Regarding the limited amount of audio transcribers and accessible multimedia analyzed, we believe that this new and complementary approach would represent a contribution to the existing literature as well as an increase in the system's applicability. Finally, in a purely experimental scenario, we plan to measure the time required for annotation, to further study the tool performance. **MM**

References

1. *Web Content Accessibility Guidelines 1.0 (WCAG 1.0),* World Wide Web Consortium (W3C), 1999; http://www.w3.org/WAI/intro/wcag.php.
2. *Web Content Accessibility Guidelines 2.0 (WCAG 2.0),* World Wide Web Consortium (W3C), 2008; http://www.w3.org/WAI/intro/wcag.php.
3. L. Moreno, P. Martínez, and B. Ruiz-Mezcua, ''Disability Standards for Multimedia on the Web,'' *IEEE MultiMedia,* vol. 15, no. 4, 2008, pp. 52-54.
4. D. Brezeale and D.J. Cook, ''Learning Video Preferences Using Visual Features and Closed Caption,'' *IEEE MultiMedia,* vol. 16, no. 3, 2009, pp. 39-47.
5. D. Bulterman et al., *Synchronized Multimedia Integration Language (SMIL 3.0) Specification,* World Wide Web Consortium (W3C), 2008; http://www.w3.org/TR/SMIL3/.

Fernando Paniagua-Martín is a teaching assistant at the Carlos III Technical University of Madrid, where his is also completing a PhD in computer science. His research interests include audiovisual accessibility on the Web and Web 2.0 and 3.0. Paniagua has an MS in computer science and technology from Universidad Carlos III de Madrid. Contact him at fernando.paniagua@uc3m.es.

Angel García-Crespo is the head of the SofLab Group in the computer science department at the Universidad Carlos III de Madrid and is the head of the Institute for Promotion of Innovation Pedro Juan de Lastanosa. His research interests include Web 2.0 and 3.0, software in relation to service, software engineering, and audiovisual accessibility. García-Crespo has a PhD in industrial engineering from the Universidad Politécnica de Madrid. Contact him at angel.garcia@uc3m.es.

Ricardo Colomo-Palacios is an associate professor in 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. Colomo-Palacios has a PhD in computer science from the Universidad Politécnica of Madrid. Contact him at ricardo.colomo@uc3m.es.

Belen Ruiz-Mezcua is a lecturer in the computer science department of the Universidad Carlos III de Madrid and is vice-chancellor of research adjunct to the scientific park. Her research interests include biometrics and speech processing and artificial intelligence. Ruiz-Mezcua has a PhD in physics from the Telecommunications School of Polytechnic University of Madrid. Contact her at mbelen.ruiz@uc3m.es.