A Multimedia System to Produce and Deliver Video Fragments on Demand on Parliamentary Websites

Elena Sánchez-Nielsen[1], Francisco Chávez-Gutiérrez [1,2], Javier Lorenzo-Navarro[3], Modesto Castrillón-Santana[3]

[1]*Departamento de Ingeniería Informática y de Sistemas – Universidad de La Laguna, 38271 Santa Cruz de Tenerife, Spain*
enielsen@ull.edu.es
[2]*Parlamento de Canarias, Teobaldo Power 7, 38002 Santa Cruz de Tenerife, Spain*
fchavez@parcan.es
[3]*SIANI – Universidad de Las Palmas de Gran Canaria, 35017 Gran Canaria, Spain*
{javier.lorenzo, modesto.castrillon}@ulpgc.es

Abstract

Parliamentary websites have become one of the most important windows for citizens and media to follow the activities of their legislatures and to hold parliaments to account. Therefore, most parliamentary institutions aim to provide new multimedia solutions capable of displaying video fragments on demand on plenary activities. This paper presents a multimedia system for parliamentary institutions to produce video fragments on demand through a website with linked information and public feedback that helps to explain the content shown in these fragments. A prototype implementation has been developed for the Canary Islands Parliament (Spain) and shows how traditional parliamentary streaming systems can be enhanced by the use of semantics and computer vision for video analytics. The semantic web technologies used make search capabilities on parliamentary websites available to users to retrieve video fragments on demand with accurate and timely information. In addition, video analytic techniques enable the automation of identifying representative keyframes to be annotated by parliamentary experts. As a result, parliaments are able to enhance citizens' access to information and ensure that these institutions are more open and accountable on their websites and; at the same time, the labor-intensive tasks of parliamentary experts are considerably reduced.

Keywords: multimedia system; video fragments on demand; semantic web technologies; video analytics; person re-identification

1 Introduction

Since 2000, parliamentary websites have become one of the most important windows to view the work of legislatures and to hold parliaments to account. During the past decade the goals of these websites have become more complex and challenging, and parliaments around the world aim to improve the access to their content using multimedia technologies [25]. One of the most recent challenges has come from providing viewing video fragments on demand on parliamentary activities (plenary sessions and committee activities). With the term "video fragments on demand" we refer to watching a specific fragment for a specific information need from the video streaming service when users choose to, e.g. watch a video clip of 15 minutes on the latest health policy reform within a three-hour video streaming 14 days after the broadcast time. Therefore, it has become a crucial aim for most parliamentary institutions to provide new multimedia solutions capable of displaying video fragments on demand on plenary activities, and linking relevant information that helps to explain the content shown in these fragments, given the fact that most working-age citizens are unable to view streaming live and, at the same time, parliamentary streaming videos are time intensive.

Nowadays, 52% of parliamentary institutions around the world are making live streaming video available through their websites to provide access to their plenary meetings and another 37% of parliaments have been considering developing this capability [21]. However, exploiting the full potential of tools that enable users to deal with videos to find what they are looking for quickly, and get all the related parliamentary information and public feedback is to date still limited. Up till now, very few parliaments (e.g. the U.S. House of Representatives and the European Parliament) are providing basic capabilities that add to their streaming video, such as linking the video recording and the text of statements and linking the bills and votes shown in the video. However, none of them are providing video fragments on demand and linking these fragments to relevant parliamentary information and public feedback [21].

In this paper, we present a multimedia system for parliamentary institutions to produce and deliver video fragments on demand through a parliamentary website. The key issues we address are: (1) how to describe and annotate plenary and committee videos in order to make available search capabilities to retrieve video fragments on demand, (2) how to use the annotations to provide at the user interface the capability of displaying in a simultaneous way parliamentary information and public feedback on the legislative initiative being debated and accurate information on their representative members who are intervening whilst users are watching the content of the retrieved fragments, and (3) how to automate the process of identifying the key representative frames to be annotated.

The Canary Islands Parliament, Spain (http://www.parcan.es/) has collaborated with us in the development of a prototype implementation of this multimedia system with two purposes: (1) to improve the current streaming system (http://www.parcan.es/video/canales.py) and, (2) to enhance citizens' access to parliamentary information on its parliamentary website in order to be more open, transparent and accountable. The contributions of our approach are threefold: (1) the presentation of an overall integrated architecture, including the description of a semantic model by following linked data principles to represent all the important entities and relationships for the plenary meetings domain; (2) automatic identification of speakers in plenary videos as a visual re-identification problem to determine speakers' appearances over a camera network, to subsequently define what keyframes need to be annotated; and (3) a

user interface on the parliamentary website to make video fragments available on demand with accurate and well-timed information and public feedback.

The rest of the paper is organized as follows. Section 2 presents the related work on the different aspects of the development and implementation of our multimedia system. Section 3 describes the overall integrated architecture solution and our approach to address the three key issues described above. Section 4 presents the experimental results for the re-identification problem and, Section 5 highlights the conclusions and future work.

2 Related Work

2.1 Video fragments on demand retrieval problem

Retrieving video fragments on demand is a complex task due to the temporal nature of videos. This retrieval problem requires tackling two main issues: how to address fragments in videos and how to describe these video fragments [23].

2.1.1 Addressing video fragments

The methods for addressing video fragments are focused on Uniform Resource Identifier (URI) and non-URI based mechanisms. In the non-URI based class of solutions, SMIL [11], MPEG-7 [26], and TV-Anytime [19] are the most representative specifications. The former standard allows playing a temporal fragment of a video; however, it entails a huge waste of bandwidth when large videos are used because the complete media resource has to be obtained [35]. Using MPEG-7, a video can be decomposed into *VideoSegments* that can be described by a *MediaTimePoint* and *MediaDuration* corresponding to the starting time and media duration respectively. Using TV-Anytime [19], temporal intervals can also be defined and accessed through *segments* within a video. In the URI-based solutions, the MPEG-21 [27] specifies a normative URI fragment syntax to address fragments in MPEG compatible files, but the supported media types are restricted to MPEG formats only. In recent years, a new specification has been provided by the W3C Media Fragment Working Group [51] referred to as the Media Fragment URI 1.0 specification (MFURI 1.0) [52]. This specification [35] offers a standard means of addressing media resources (video, image and audio) via four dimensions (temporal, spatial, track and named section) using URIs.

2.1.2 Describing video fragments

The aim of describing video fragments is to enable annotations on video fragments. So, instead of searching for video fragments directly, a user is able to search for the annotations that have been added to them. The approaches to tackle these annotations can be classified according to standards prior to the paradigm of the semantic web and approaches based on this paradigm.

The semantic web [4] is an evolving development of the World Wide Web, in which the meaning of the information on the web is defined. The basic idea is to use ontologies to accurately describe contents in a machine readable way. Ontologies define common, shareable and reusable views of a domain, and they give meaning to information structures that are exchanged by information systems [42]. Before the rise of the semantic web, MPEG-7 [40] and MPEG-21 [27]

were the most representative standards to annotate video data. Nowadays, one of the most recent directions of research related to describing video fragments consists of interlinking annotations by different users (i.e. to specify links and annotations between video fragments) based on the use of the linked data paradigm [23, 32, 33, 39]. This paradigm is the revolutionary development of the semantic web and refers to a set of best practices for publishing and interlinking structured data on the Web [5]. The technical basis consists of organizing the knowledge-based resources with the following principles [8], which are known as the linked data principles:

1. Use URIs to identify things. This principle advocates using URI references to identify Web documents, world objects and abstract concepts.

2. Use HTTP URIs so these names can be looked up. This principle advocates the use of HTTP URIs to identify objects and abstract concepts, enabling these URIS to be *dereferenced* over the HTTP protocol into a description of the identified object or concept.

3. Return useful information having looked up these URIs, using semantic web standards. This principle advocates the use of a single data model for publishing structured data using the Resource Description Framework (RDF) [54].

4. Include links to other URIs, so they can discover more things. This principle advocates the use of hyperlinks to connect not only Web documents. For example, a hyperlink may be set between a person and a company.

As linked data builds directly on Web architecture, the Web architecture term *resource* is used to refer to the *things of interest*, which are, in turn, identified by HTTP URIs. In summary, linked data principles are built upon standards from the World Wide Web Consortium based on the Resource Description Framework (RDF) [54] for representing data about resources. It allows us to formulate *statements*, each of which consists of a subject, a predicate, and an object. The RDF Vocabulary Description Language, also called RDF Schema (RDFS) [55], and the Web Ontology Language (OWL) [49] are used to describe the terms, classes, properties and relationships used in a RDF model. An RDF store can be queried via the SPARQL Query Language for RDF datasets [50] through a SPARQL endpoint. The most promising dataset of linked data is the Linked Open Data (LOD) cloud [24].

Up till now, a reduced number of multimedia applications have been developed using linked data principles fundamentally due to a lack of awareness and understanding between multimedia and semantic communities [39]. In the parliamentary context, a recent trial has highlighted the basic research questions when linked data is applied to media fragments and annotations, and has shown how a specific video fragment is linked to simple annotations such as speaker's information in the LOD cloud for the UK Parliament [33]. In this paper, our approach takes a different direction as far usage of linked data principles for fragments of multimedia videos are concerned in [23, 32, 33]. We apply the linked data principles as modeling principles to explicitly identify each parliamentary resource, linking these resources to each other and to external resources and, enriching our domain ontology from the LOD cloud. It provides us with the foundation for annotating the different resources existing in the plenary meetings domain and link these annotations to precise video fragments based on both time instance and duration of plenary videos.

2.2 Segmentation of speakers' speeches

The aim of segmentation of speakers in the context of parliamentary institutions is to identify the set of segments within a plenary video that corresponds to each speaker's speech. Given the elevated quantities of video data, such a task is generally labor-intensive. In addition, the results of this task are the basis on which parliamentary experts make their annotations on plenary activities.

The speakers' segmentation problem has been traditionally formulated as an automatic labeling approach using speech features [3, 36]. However, in recent years, the computer vision community has shown the potential of using visual features, outperforming audio based systems [28]. In a computer vision context, this problem can be seen as a re-identification issue. It involves determining whether an identity of interest has previously been observed [12]. To automatically match people's identities over time, a general re-identification task typically takes the following steps: (i) extracting concise imagery features, (ii) constructing a representation (model) capable of both describing and discriminating individuals, and (3) matching specified probe images or tracks against a gallery of persons in previous camera views by measuring the similarity between the images. Most computer vision identity modeling approaches are focused on face patterns or facial descriptors [30]. It should be noticed that in most video scenarios other more robust biometric traits (e.g. iris) are not available. Therefore, a robust modeling of the global appearance of an individual (e.g. clothing, gender, and hair) is necessary for re-identification.

2.3 Feedback in parliamentary institutions

Providing the feedback from public opinion on parliamentary activities is an essential aim for parliamentary institutions to foster openness, transparency and participation [21]. Twitter is currently the most popular microblogging service used to communicate with parliaments [6, 21]. Specifically, facilitating individual twitter posts on retrieved video fragments on demand enable users to know personal opinions, concerns, and preferences on the specific topics of these retrieved fragments. Individual twitter posts can be represented in the microblogging domain using the following ontologies from the LOD cloud: FOAF [10] to model user profiles, SIOC [9] to describe microblog entries, and OPO [41] for modeling the semantics of a user's appearance in the online world. On the other hand, the application of sentiment analysis on twitter posts can help users to know if the polarity of posts tends towards positive, negative or neutral [29, 38].

3 Multimedia System: architecture, functionalities and implementation

3.1 System architecture

The multimedia system to produce and deliver video fragments on demand is designed around the conceptualization of the plenary meetings domain in the Canary Islands Parliament. A domain ontology was developed using linked data principles to provide the concepts and relationships of the different entities that exist for the domain of plenary sessions and committee meetings. The multimedia system is based on the architecture shown in Fig. 1. The system architecture consists of three layers: data, production, and delivery.
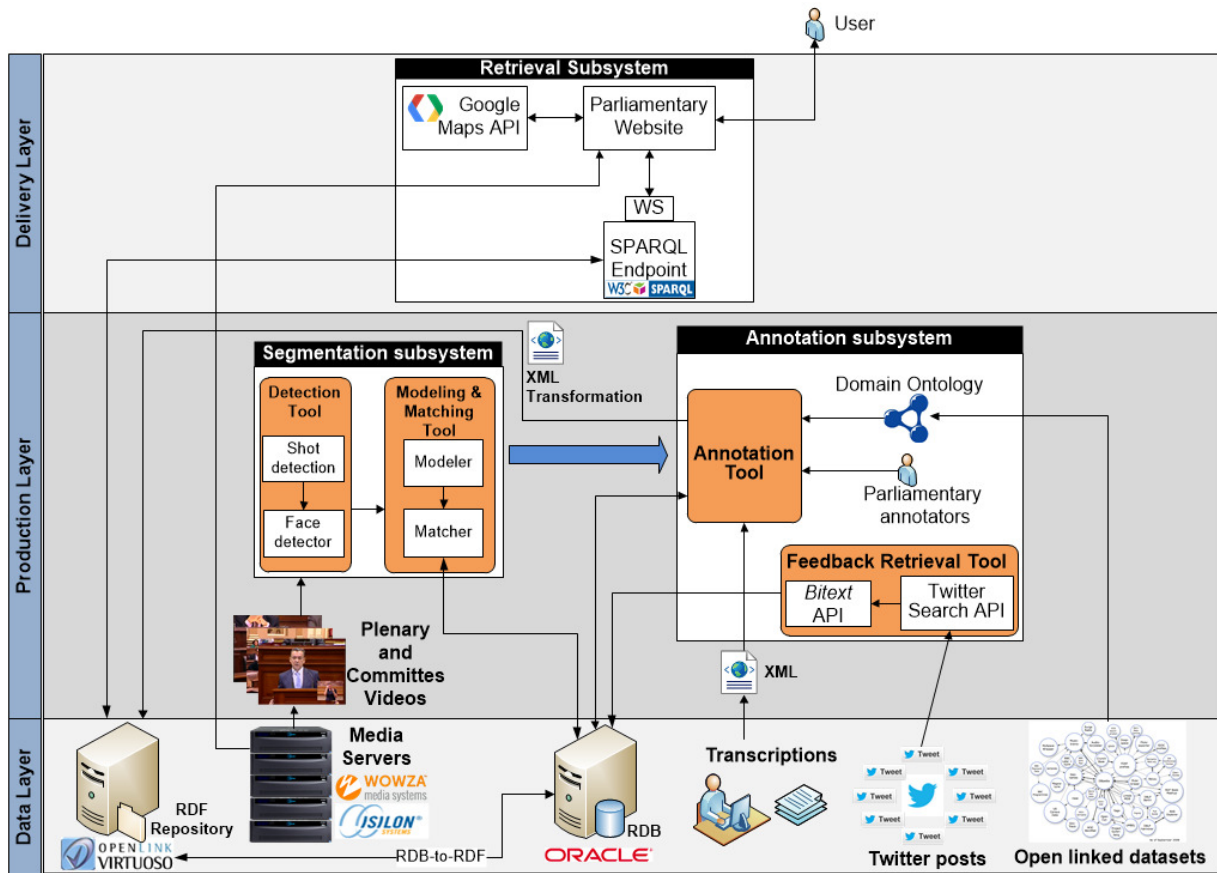
Fig. 1 Multimedia system architecture

The bottom data layer is composed of a RDF repository [37] that contains the RDF triples on parliamentary debates and committee meetings, the media servers to store and stream on-demand videos, a relational database that contains the data storage for the legacy system and the new subsystems developed, XML documents with the textual transcriptions of speakers' speeches, twitter posts to track public feedback on the debate topics, and vocabularies from the LOD cloud to enrich our domain ontology. The production layer consists of:

- The segmentation subsystem which is used to automatically compute the keyframes from the plenary videos to be annotated by the parliamentary experts.
- The annotation subsystem which is employed by the parliamentary experts to annotate each separate activity taking place in a plenary session and/or committee meeting.

The delivery layer enables users with a user interface through the parliamentary website to search for video fragments for a particular information need related to the activity taking place in plenary meetings. The architecture of the multimedia system is based on three main subsystems with different related functionalities which will be described in the following sections: segmentation and annotation subsystems both corresponding to the production layer and, the retrieval subsystem corresponding to the delivery layer.
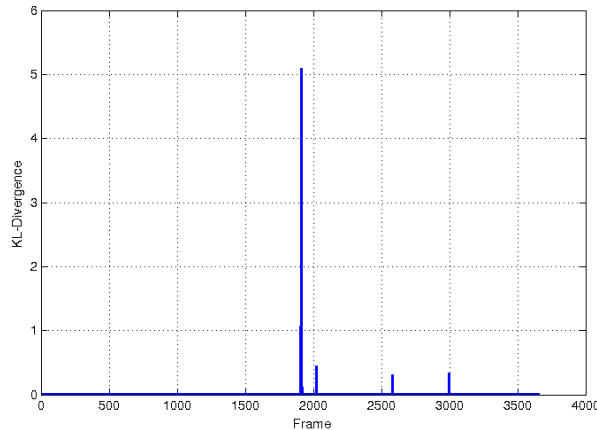
6

Fig. 2 Results of the value of the Kullback-Leibler divergence for 4000 consecutive frames. The largest value indicates a shot boundary

3.2 Segmentation subsystem

The segmentation subsystem addresses the problem of assisting the identification of the start and end points of each speaker's speech on the plenary video timeline. The set of segments computed are subsequently used by parliamentary experts as representative keyframes to be annotated using the annotation tool. We address the segmentation problem as a visual speaker re-identification issue. However, unlike traditional re-identification problems, where the aim is to determine the presence of an individual over a network of cameras; in our context it is also necessary to detect how many times the individual (speaker) appears and the temporal sequence of appearances over a camera network providing different planes of the Chamber of Deputies which include changes in pan, tilt, zoom and color adjustment during each speaker's speech. For this task, we use different visual information cues focused on classical biometrics such as face based descriptors and soft biometrics attributes of an individual speaker based on clothing and hair based descriptors. Our approach (see Fig. 1) uses a detection tool and a modeling and matching tool in order to detect the presence of a speaker, modeling it, and establishing if this detected speaker has been previously observed or if he/she is a new speaker. Our approach has the benefit of not requiring any prior model about the speakers who will intervene in the plenary meetings. In the following subsections, we describe the different steps performed by the detection, modeling and matching tools.

3.2.1 Shot detection

The issue of segmenting a video into sequences of frames sharing the same visual features, usually called shot detection, is a crucial task for our detection tool in order to be able to assign a different identity tag to each speaker who participates in the plenary meeting. Thus, the division of the video into shots is the first step to be carried out by the detection tool. The shot boundary detection problem has traditionally involved the computation of visual differences between consecutive frames and a shot has been detected when this difference has been higher than a

Fig. 3 Illustration of a shot detected using the proposed boundary detection method

certain threshold [31, 43, 46]. In our approach, the feature used in each frame to detect the shot transition is the color, which has proven to be effective [47]. Thus, each frame $frame_i$ is modeled by its color probability distribution P_i, and the shot boundary detection problem is stated as the search for a significant difference in color distribution between consecutive frames, $frame_i$ and $frame_{i+1}$. The dissimilarity of the color distribution of two consecutive frames, P_i and P_{i+1}, is measured using the Kullback-Leibler (KL) divergence [15], $D_{KL}(P_i \| P_{i+1})$. The color distribution is gathered from the histogram of a color space, which allows us to use the expression of the KL divergence for discrete probability functions:

$$D_{KL}(P_i \| P_{i+1}) = \sum_c P_i ln \frac{P_i}{P_{i+1}} \tag{1}$$

YCbCr has been used as color space and for each component a discretization process into 16 values has been applied. After that, a normalization stage is performed and finally the histograms are concatenated to make up a single histogram for each frame. Fig. 2 shows the value of the KL divergence (eq. 1) for a video where the value of the measure is close to zero for all the consecutive pair of frames, except when a shot boundary appears. Thus, the shot boundary detection problem can be stated as:

$$frame_i \text{ is boundary } \begin{cases} \text{true} & \text{if } D_{KL}(P_i \| P_{i+1}) > thr \\ \text{false} & \text{otherwise} \end{cases} \tag{2}$$

where *thr* denotes a given threshold. Fig. 3 shows a shot detected with the proposed method. It should be noticed that this shot is not a static plane because the camera makes a panning movement as can be observed in the sequence of images. The upper left and lower right images correspond respectively to the first and last frame of the shot. The remaining images are intermediate frames to illustrate the panning movement of the camera.

Fig. 4 Areas extracted from each frame to obtain the descriptors of the speaker: magenta bounding box contains the hair descriptor area, green bounding box contain the face descriptor area and the red bounding box contains the torso descriptor area

3.2.2 Face detector

For each frame in the video, if there is a speaker, he/she must be detected in order to characterize him/her to give an identity when a shot boundary is detected according to eq. (2). The plenary meeting scenario allows us to introduce some restrictions in the kind of shots that belong to the speaker's exposition. We only consider one face per frame and when there is more than one face, the largest one is assumed to belong to the speaker because he/she always will be in the foreground.

The process of face detection is carried out in two stages. In the first stage, the Viola-Jones face detector [48] is used to obtain face candidate regions. In the second stage, a verification process is performed by finding the two eyes in the candidate region making use of the Castrillon et al. eye detector approach [13].

3.2.3 Identity modeling

Once a face is detected in the frame *frame$_i$*, a set of descriptors is computed to characterize the speaker. These descriptors are extracted from three areas of the image (see Fig. 4) corresponding to: face, torso and hair location estimation.

- *Face based descriptors:* a first set of descriptors is computed from the face area (green bounding box in Fig. 4). Previously, a normalization process is carried out to rescale the face area to 48 x 48 pixels. Local descriptors such as Local Binary Patterns (LBP) and Histograms of Oriented Gradients (HOG) have been recently applied for people analysis [2, 17]. In our approach, we have used HOG because it is robust to changes in illumination and small displacements. The computation of the face descriptor is performed by dividing the face area into a 3x3 grid of cells, and the HOG descriptor is computed for each cell, $FaceHOG_{k,l}$ for $k, l \in \{1,2,3\}$. The final face descriptor, *fd$_i$*, for frame, *frame$_i$*, is the concatenation of the nine cell HOGs:

$$fd_i = \left(FaceHOG_{1,1}, FaceHOG_{1,2}, \dots, FaceHOG_{3,3}\right) \tag{3}$$

- *Hair based descriptors:* the descriptors from the hair area (magenta bounding box in Fig. 4) in the corresponding frame are obtained in a similar way to those from the face area: the area is divided into a 2x2 grid of cells where in each cell the HOG descriptor is computed, $HairHOG_{k,l} for\ k, l \in \{1, 2\}$. The division into a coarser grid is due to the fact that the hair does not have so many details as the face. The hair descriptor, hd_i, is the concatenation of the HOG of each cell in the hair area of the corresponding frame, *frame_i*:

$$hd_i = \left(HairHOG_{1,1}, HairHOG_{1,2}, \dots, HairHOG_{2,2}\right) \tag{4}$$

- *Clothes based descriptors:* the torso area (red bounding box in Fig. 4) covers part of the upper clothes of the speaker. This information, along with face and hair, can be used to better discriminate among speakers particularly in situations with large face pose changes [43]. In a plenary meeting scenario, people normally wear the same clothes during the whole session so this visual cue can give us valuable information about the speaker. In order to characterize the clothes, color information is used because it has been shown to give good results in clothes segmentation problem [34]. Therefore, the descriptor chosen for the clothes is the 2D histogram of the YCbCr components, with each component discretized into 16 different values. The use of a unique histogram presents some limitations in the plenary meeting scenario given the cultural fact that men will be commonly wearing a similar dark suit with only some variations in the shirt and necktie color. To make our approach more robust, three different histograms are computed: one central one (*histogram_central*) covering the shirt and necktie and the other two (*histogram_left* and *histogram_right*) covering the suit. Thus, the descriptor for the clothes, *cd_i*, is composed of the concatenation of the three histograms after normalizing them:

$$cd_i = \left(histogram_{left}, histogram_{central}, histogram_{right}\right) \tag{5}$$

3.2.4 Speaker matching

After the speaker descriptors have been computed for frame *frame_i*, a dissimilarity measure is defined to match the current speaker of *frame_i*, *speaker_i*, with each previous speaker, *speaker_k*, in order to determine if the current speaker has been previously detected or if he/she is a new speaker. Due to the different nature of the visual information sources used (face, hair, and clothes), a combination of three dissimilarity measures is defined. On the one hand, the dissimilarity of the face and hair (*D_face* and *D_hair*) is measured by means of the cosine distance:

$$D_{face}\ (speaker_i, speaker_k)\ = \text{cosine } distance(fd_i, fd_k) \tag{6}$$

$$D_{hair}\ (speaker_i, speaker_k)\ = \text{cosine } distance(hd_i, hd_k) \tag{7}$$

where fd_i and hd_i correspond respectively to the face and hair descriptors (eq. 3 and eq. 4) of $speaker_i$ in $frame_i$ while fd_k and hd_k correspond to those descriptors of any speaker previously detected. The dissimilarity measure of the clothes ($D_{clothes}$) is computed using the KL divergence between the color histograms as follows:

$$D_{clothes}(speaker_i, speaker_k) = \sum_c cd_i \ln \frac{cd_i}{cd_k} \qquad (8)$$

where cd_i denotes the clothes' color descriptor (eq. 5) of $speaker_i$ and cd_k denotes the clothes color descriptor of any speaker previously detected. The process of identity assignation to the detected speaker in $frame_i$ using the three dissimilarity measures is as follows:

$$speaker_i \text{ identity} = \begin{cases} k & \text{if } D_{clothes}(speaker_i, speaker_k) < c_{thr} \text{ and} \\ & D_{face}(speaker_i, speaker_k) < f_{thr} \qquad \text{and} \\ & D_{hair}(speaker_i, speaker_k) < h_{thr} \\ \text{new id} & \text{otherwise} \end{cases} \qquad (9)$$

where c_{thr}, f_{thr} and h_{thr} denote predefined threshold values. The matching process defined in (eq. 9) is performed for each detected speaker, assigning a new identity, named *id*, if there is no match with previous speakers. Finally, each time a shot boundary is detected, an *id* label is assigned to the just concluded whole shot. This is computed based on a majority voting strategy that assigns to the shot the *id* label of the most frequent speaker.

3.3 Annotation subsystem

The annotation subsystem aims to provide an accurate description of the parliamentary activity taking place in plenary meetings. This subsystem consists of two key tools:

- *The Annotation Tool:* this tool enables parliamentary experts to annotate each agenda item taking place in a plenary session and/or committee meeting in the form of a plenary video timeline with respect to the domain ontology.
- *The Feedback Retrieval Tool:* a hashtag is provided for each relevant legislative initiative to be dealt with on the plenary agenda of a specific plenary meeting. The feedback retrieval tool computes the dataset of twitter posts that have been tagged with specific hashtags with the Twitter search API [44]. Subsequently, the Bitext API [7] is used to process sentimental analysis on each twitter post, so each public feedback made by a user can be retrieved with a score between -1.0 and 1.0; where -1.0, 0 and 1.0 denotes respectively negative, neutral and positive opinion from public.

In the following subsections, first, the structure of plenary meetings is introduced. Second, the semantic modeling principles proposed to model the conceptualization of the plenary domain are described. Third, the domain ontology is presented. Fourth, the annotation tool to annotate the parliamentary activity is described. Fifth, examples of RDF graphs for a plenary session are presented in order to show the backbone of the proposed approach.

3.3.1 Plenary sessions and committee meetings

In a parliamentary context, a plenary meeting is a forum that is attended by the appointed members of different political groups following a defined structure according to an agenda. This agenda consists of a set of topics, most of which are legislative initiatives. Each item on the agenda is referred to as an *agenda item*. All initiatives have a proponent and are firstly presented by the chairman. Secondly, a number of speakers intervene in a chronological order. Finally, when the debate is finished, decision making is applied to the legislative initiative. This decision making can lead or not to voting.

3.3.2 Semantic modeling

We adopt linked data principles as basic principles of the semantic modeling of our plenary meetings domain in the following way:

1. Use of URIs as identifiers which can be dereferenced using HTTP protocol (to follow the first and second rule of linked data principles). It means that each resource of a plenary meeting (agenda item, legislative initiative, member, intervention, transcription, and twitter post) is unique and clearly identified.
2. Data on plenary meetings resources are represented using the standard RDF (to follow the third rule of linked data principles), and therefore, SPARQL queries can be carried out on a RDF repository to retrieve information on specific video fragments and related parliamentary data.
3. Use of links to other URIS (to follow the fourth rule of linked data principles) in order to connect data inside the parliamentary institution, as well as to connecting parliamentary resources with external resources (LOD cloud and external government information systems).

In a nutshell, linked data principles provide us with the potential to reuse ontologies from the LOD cloud to enrich our domain ontology; the ability to explicitly identify each parliamentary resource, which ensures consistent reference and linking; and the capability of linking these resources with each other and with external resources.

3.3.3 Domain ontology

An ontology was developed to model the different entities and relationships that exist for the domain of plenary meetings related to members of the parliament, legislative initiatives, and activity taking place in these meetings. The ontology also relates each separate activity to precise video fragments. The most relevant concepts and relationships related to the domain ontology are described below.

- *Members of parliament:* the domain ontology provides the concepts and relationships between the members of the parliament and the proposed legislative initiatives, speeches, the role they play in the parliament, parliamentary body, and committee membership. Two main vocabularies are reused from the LOD cloud: *Friend of a Friend* (FOAF) [10], and *GeoNames* [45] to represent the member of parliament as a person and add geospatial information about the constituency of each member. The domain ontology links the members of parliament with different parliamentary resources through the definition of the following specific

12

properties: *hasLegislature* (to link members to a legislative mandate), *isMemberOf* (to link members to committees), *belongsTo* (to link members to a parliamentary body), *hasRole* (to link members to the role they play in the parliament), *isProposedBy* (to link members to a proposed legislative initiative).

-   *Legislative initiatives:* the domain ontology provides the concepts and relationships between the types of legislative initiatives and the specific procedure to process each type of initiative. Nine different *subclasses* of legislative initiatives are defined: legislative, normative, control on government actions, promotion of government actions, actions related to unconstitutionality matters, actions related to commissions, actions ated to institutions and organizations, and actions related to the parliament as an organization. Each one of the different legislative initiatives is related to their corresponding procedural actions through different specific properties. The Dublin Core vocabulary [18] is reused to identify the *title* and *identifier* of a legislative initiative.

-   *Parliamentary activity:* each separate activity taking place in a meeting is represented by a specific item on the meeting agenda. It is modeled as an event in the domain ontology characterized by: *what happened*, *who* was involved, *where it happened*, and *when it happened*. The feature *what happened* identifies each legislative initiative/topic to be debated, the public feedback on this initiative available as twitter posts, and the decision making adopted for this initiative after the debate has finished. This event aspect is characterized by the properties: *hasItem*, *hasHashtag*, *hasDecision* and *hasAssociatedTweet*. The properties *hasGeolocation* and *hasSentiment* are respectively used to relate a twitter post to its geolocalisation and its sentimental analysis (1.0: positive; 0: neutral; and -1.0: negative). The ontology SIOC [9] is reused to represent the topic, the user, the date, and the content. The feature of *who* was involved is structured around different speakers' interventions on the agenda item to be debated. The event aspects *where it happened* and *when it happened* are modeled reusing the properties from the Dublin Core *dc:place*, *dc:date*, *dc:title*, and *dc:language*. The whole plenary meeting is related to a video recording resource, and different speakers' speeches related to each legislative initiative to be debated are related to video fragments according to the temporal instants and durations on the video timeline. The property *hasVideo* is used to relate a plenary meeting resource URI to a video recording resource URI whilst the property *hasFragment* is employed to relate a speaker speech to a video fragment. The Ontology for Media Resources [53] and the Timeline ontology [1] are reused respectively to describe the core properties of plenary video resources and to identify the temporal start and duration on the video timeline corresponding to different video fragments.

3.3.4 Annotation tool

Each plenary meeting is annotated by the parliamentary experts according to the two first aspects that model an event in the domain ontology: place and date. Each plenary meeting agenda item is annotated according to the other two aspects that model an event. Keywords related to the agenda item are also annotated by the parliamentary experts using the Eurovoc [20], which is a multidisciplinary ontology-based thesaurus covering the activities of the EU, and
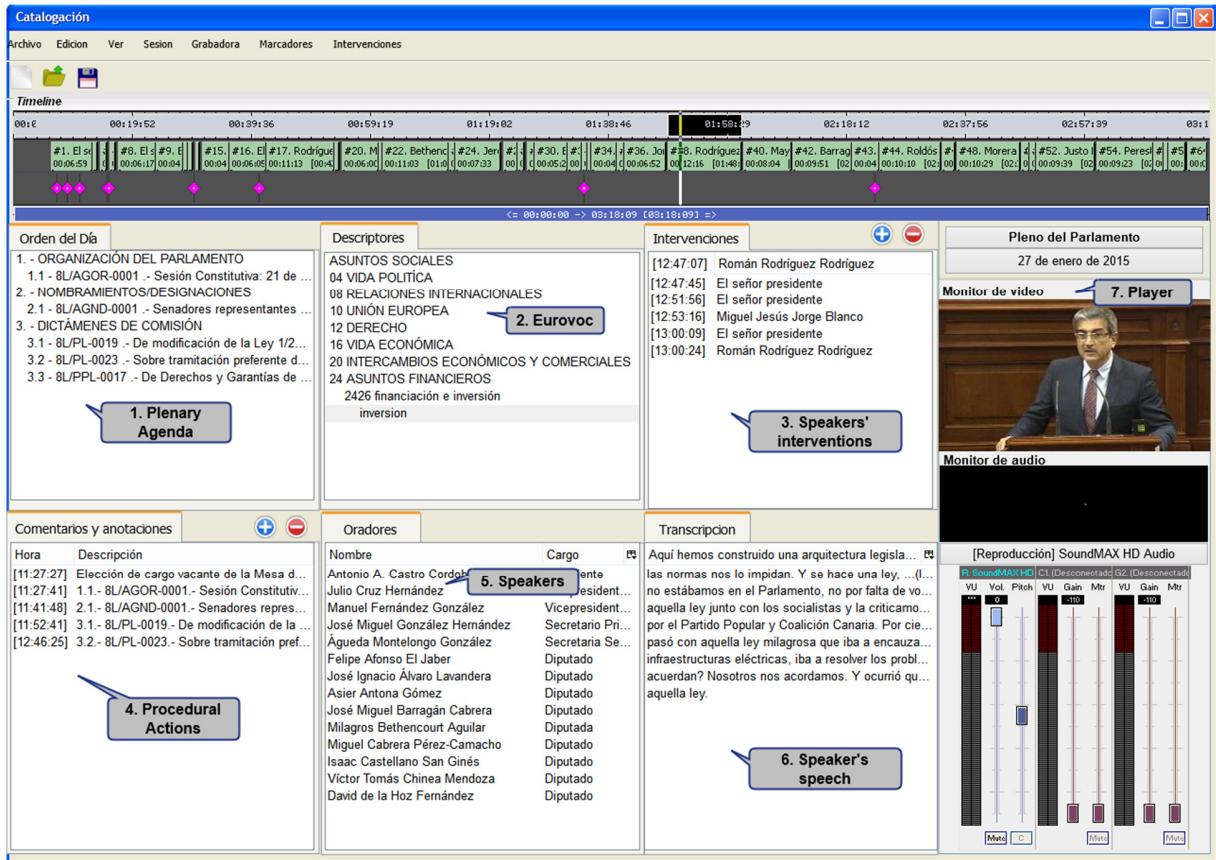
13

Fig. 5 A screenshot for the annotation tool

in particular the European Parliament. For each plenary meeting to be annotated, the annotation tool receives the following input files:

- The start and end points of each speaker's speech generated by the segmentation subsystem as keyframes to be annotated.
- The text transcript of the different speakers interventions as a XML file.

Using the annotation tool, the parliamentary experts associate each speaker's speech with the corresponding XML tag that contains the text based transcription of the speech. The output of this process is a transformed XML file that contains text transcriptions of speakers' interventions linked to the corresponding speakers fragments. The annotation data that describe speakers' speeches are linked to the corresponding plenary meeting video fragments.

The tool interface for parliamentary experts consists of seven sections (see Fig. 5): (1) Agenda: includes the items to be treated in the plenary meeting; (2) Keywords: provides the relevant keywords from the Eurovoc ontology-based thesaurus to describe the legislative initiatives; (3) Speakers' intervention: provides the temporal starts and ends of each speaker's speech. The timeline corresponding to the different speakers to be annotated is supplied by the segmentation subsystem; (4) Procedural actions: the procedural actions and the decision making adopted for the

14

legislative initiative are added according to the domain ontology; (5) Speakers: shows the list of available speakers in a given legislature; (6) Speakers' speech: provides the text transcriptions of the different speakers' speeches; and (7) Video player: displays the video according to the timeline.

Fig. 6 and 7 show an RDF graph representing the data according to the domain ontology for a plenary session held on 27th January, 2015 on a law proposal (8L/PL-0023 on preferred processing of strategic investments for the Canary Islands). The domain ontology is referred to as in these figures to as *opc*.

The backbone of the plenary session is illustrated in Fig. 7 and consists of a set of resources with their own unique identifiers: the session, the agenda item, the legislative initiative corresponding to this agenda item, a twitter post associated to this initiative with the sentimental analysis result, and the video recording of this session. The domain ontology specific property *hasItem* is attached to the resource containing the plenary session URI and points to the range of possible agenda items, whilst the property *hasSpeechIntervention* is attached to the resource containing an agenda item URI and points to the range of possible interventions that the legislative initiative as a whole can contain. The specific property *Subject* relates the agenda item URI to the legislative initiative URI. This initiative is related to the proponent by the specific property *isProposedBy*. The property *hasHashtag* relates the initiative to the specific hashtag used for twitter for this initiative (in this case, pcPL23). Every twitter post is related to the agenda item by the property *hasAssociatedTweet* and to the sentimental analysis score by the property *hasSentiment*. The property *hasDecision* relates the decision making adopted for the agenda item with the literal value *approved* after the debate has finished, and the property *hasVoting* relates the agenda item to the number of votes.

The data about the second intervention on the legislative initiative being discussed, the speaker and the video fragment associated with this intervention are illustrated in Fig. 7. Each intervention is related to the specific speaker by the property *hasActor*. The speaker is connected to his text transcription of his speech through the property *hasTranscription* whilst the property *hasRole* relates the legislative initiative to the role played by the speaker in the debate (in this case, representative member of the Canary Islands Parliament). The properties *isMemberOf*, *belongsTo*, and *hasLegislature* represent the political background information and relate the speaker to committee, parliamentary body and legislative mandate, respectively. The property *hasFragment* relates the intervention of the speaker to a fragment video, which is described by reusing the Media Resources and Timeline ontologies [53, 1]. The speaker is linked to his entry in DBpedia [16] with the property *owl:sameAS*, whilst the property *rdfs:seeAlso* is used to link the URI of the speaker to information from Wikipedia about him.

3.3.5 Conversion

The W3C standard RDF is the standard data model in which data on plenary meetings resources have to be available according to the third linked data principle. Therefore, we have taken the data sources stored in the parliamentary relational database related to the parliamentary activity annotations, representative members, legislative initiatives, twitter posts associated with legislative initiatives, and sentimental analysis results on each twitter post and, transform all of them into RDF. Converting these relational data into RDF will enable them to be queried through a SPARQL query engine. The Virtuoso universal server [37] is used as the tool to perform the RDB-to RDF process by using the RDF Views functionality.

3.4 Retrieval subsystem

The user interface of the retrieval subsystem is adapted to the parliamentary website of the Canary Islands Parliament, where a user who wants to get information from plenary sessions and/or committee meetings is able to express his/her information needs according to the terms "Speaker", "Initiative", "Eurovoc descriptors", "Procedural actions", "Date" and "Legislature". The visualization results for a search query with these concepts include six sections (see Fig. 8):

- Video fragment: displays the exact video fragments (e.g. a five-minute fragment within a four-hour video).
- Parliamentary information: displays background information related to the video fragments retrieved divided into three sections: legislative initiative, speaker and transcription. The first section provides information on the initiative being displayed: type of initiative, code, title, date, proponent, current status, a link to the procedural actions undertaken to date, a link to related initiatives, a link to the decision-making procedure adopted and a link to voting related information. The second section provides the legislative mandates, committee and parliamentary body membership, legislative initiatives proposed and interventions. The third section shows the transcription document.
- Twitter posts: enables users to find out the public opinion from the twitter posts related to the legislative initiative debated.
- Statistics: enables users to find out the percentage of twitter posts whose polarity tends to be positive, negative or neutral.
- Public opinion geolocalization: the Google map API is used to give users a geographical image to allow them to know the geolocalization of different opinions, obtaining an overview of the sources of twitter posts.
- Keywords: shows the most relevant topics according to the Eurovoc ontology-based thesaurus.

3.4.1 Implementation

The search results are returned by the retrieval subsystem in the delivery layer (see Fig.1), retrieving the video resources fragments from the media servers and the parliamentary information and public feedback data from the Virtuoso server. The MFURI specification [52] is used to identify video fragments and the Virtuoso server to map the relational data to RDF with Virtuoso's RDF Views functionality. Access to the RDF Views is implemented in the delivery layer using the Web services functionality provided by Virtuoso, which take as input data the users' search form and build the SPARQL queries defined for each Web service. It provides the mechanism to send SPARQL queries on the RDF data. The user interface of the parliamentary website has been developed with python programming language because the official parliamentary website of the Canary Islands Parliament is developed with this programming language on the Internet Information Service (IIS) Web server. Data returned are provided in JSON format and transformed to HTML format for results visualization.
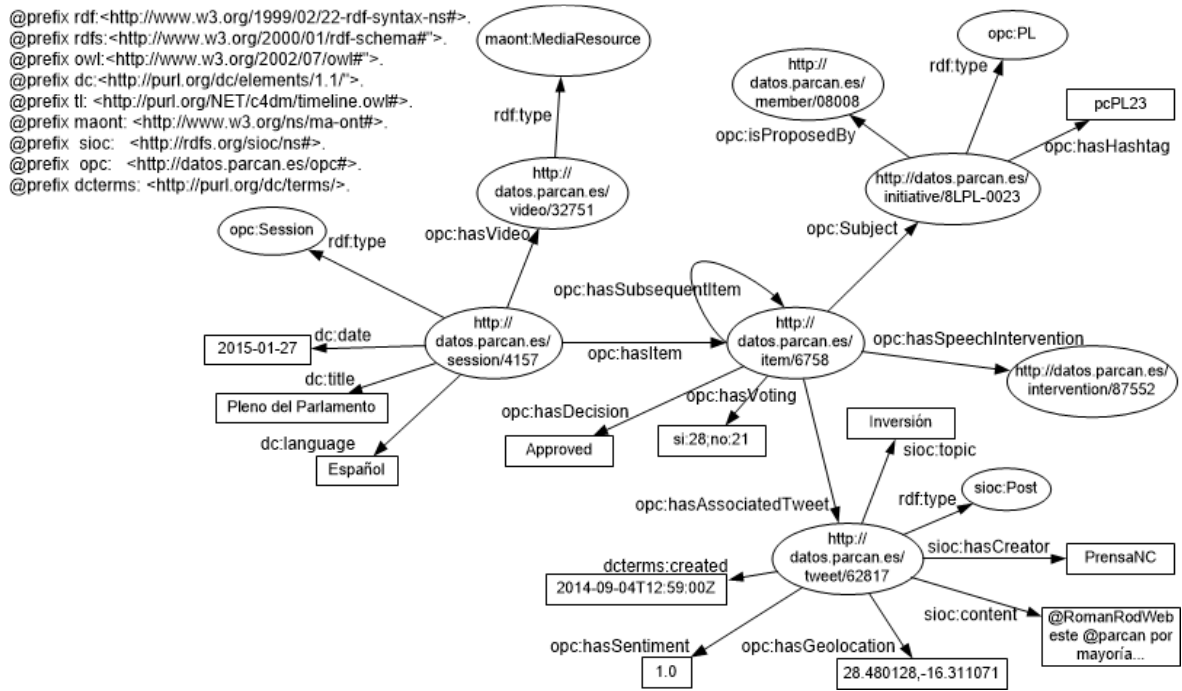
Fig. 6 RDF graph with the backbone of the model for a plenary session (27th January, 2015) on the initiative 8L/PL-0023
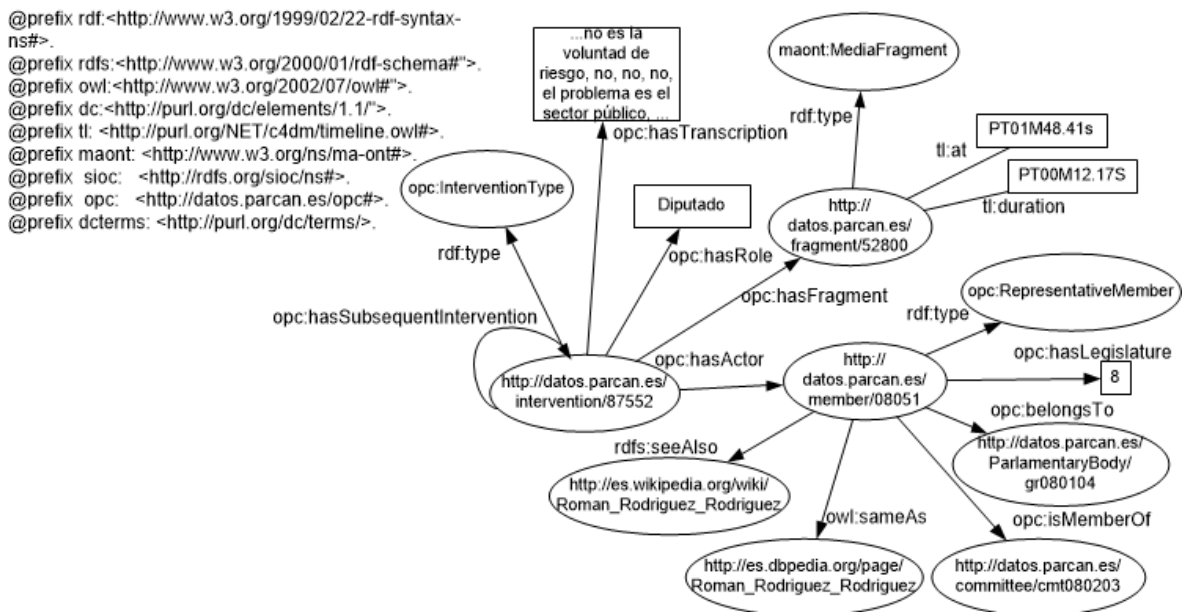

Fig. 7 RDF graph with the representation of an intervention regarding the legislative initiative 8L/PL-0023, speaker representation defined by legislature, role, parliamentary body, membership of committees, and related fragment video to this speaker intervention in the plenary session held on 27th January, 2015
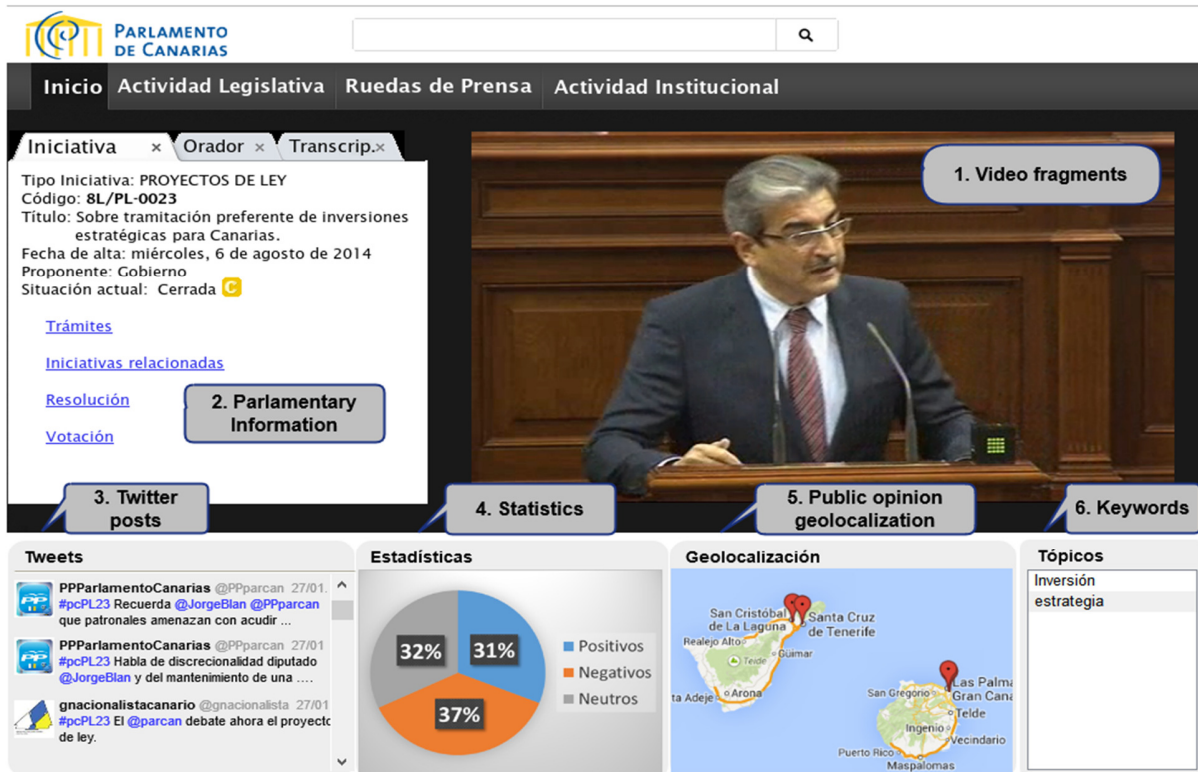
Fig. 8 A screenshot of the results for a user query on the parliamentary website

4. Experimental results

This section presents the experimental evaluation of the segmentation subsystem described in section 3.2, whose aim is to detect the presence and temporal appearances of all speakers captured by the camera network of the Chamber of Deputies. For this purpose, we have collected a dataset making use of the recording and production system implemented since September 2014 to provide live coverage of plenary sessions of the Canary Islands Parliament (http://www.parcan.es/video/canales.py). The video dataset consists of 32 videos containing a total of almost 9 million frames; resulting in approximately 12.000 shot cuts detected with the method described in section 3.2.1. An average of 19 different speakers have been recorded in each plenary session for a playback time of over 100 hours. This dataset is challenging given that the recording does not provide a single field of vision focused on each speaker's intervention. Instead of it, several and different views of the Chamber of Deputies are captured by the camera network including different individuals and changes in pan, tilt and zoom. This dataset is also characterized by clothing similarities among deputies, changing lighting conditions and automatic color adjustment during speakers' speeches, viewpoint variations across camera views when a speaker is giving the speech, speakers looking down reading the speech in most shots, cluttered background and occlusions. Fig. 9 illustrates one of the videos of a plenary session used in the experimental evaluation. The first row shows different views of a specific speaker, while the second row shows different vision fields of the Chamber of Deputies.

Fig. 9 Illustration of a plenary session video with different vision fields of the appearances of a specific speaker and the Chamber of Deputies

Since the focus of the Canary Islands Parliament is based on annotating in an automatic way the different speakers who intervene in plenary meetings for evaluation, the Parliament only provides us with a coarse-grained ground truth at the speaker level, i.e., annotations with the time when each speaker starts and ends his/her speech but no annotations when the network of cameras focuses on different views of the Chamber of Deputies during the speech. Therefore, we will focus on providing experimental results on the correct detection of the different speakers who intervene in plenary meetings.

The experimental evaluation of people re-identification techniques [46] is traditionally carried out using two sets of images/videos, named probe and gallery, corresponding to a network of cameras situated in different locations, and where each individual appears at least once in each set. Our scenario differs from the classical one in two ways. On the one hand, each video is a re-identification problem itself. On the other hand, for each individual a first decision has to be done in order to decide if he/she is a new speaker or a previously seen speaker by the system. In addition, a detected speaker cannot appear again because he/she participates only once time in the plenary session. Therefore, the widely used probe/gallery partition cannot be used in our context.

Although it was pointed out that the problem faced in this paper has not the traditional probe and gallery sets, we adopt the different measurements proposed in Cong et al. work for evaluating re-identification problems in surveillance systems [14], in order to have a reference with other re-identification proposals. These measurements adapted to our plenary session context are defined as follows:

- True re-identification (TR), also known as true match or true positive: a true re-identification occurs when shots that have the same *id* label belong to the same speaker.
- True distinction (TD), also known as true non-match or true negative: a true distinction occurs when shots that have different *id* labels belong to different speakers
- False re-identification (FR), also known as false positive or false match: a false re-identification occurs when shots with the same *id* labels belong to different speakers.
- False distinction (FD), also known as false negative or false non-match: a false distinction occurs when shots that have different *id* labels belong to the same speaker.

19

Table 1 Measures re-identification results for the evaluated dataset

Video	Length	Frames	Shots	Speakers	TRR (%)	FRR (%)	TDR (%)	FDR (%)
		Features				Measures results		
2770	3:29:22	314050	660	8	82.1	17.9	86.5	13.5
2771	0:33:23	50075	101	6	100.0	0.0	100.0	0.0
2785	2:41:54	242850	325	32	86.6	13.4	96.6	3.4
2786	2:57:00	265500	396	17	78.4	21.6	90.5	9.5
2787	5:09:20	464000	738	24	76.8	23.2	91.0	9.1
2789	2:34:54	232350	334	26	85.5	14.5	96.9	3.1
2790	2:42:18	243450	451	13	89.1	10.9	90.9	9.1
2791	4:55:05	442625	636	25	61.7	38.3	91.2	8.8
2792	1:48:00	162000	318	11	82.9	17.1	72.1	27.9
2799	2:41:17	241925	269	33	60.9	39.1	96.9	3.1
2800	3:02:12	273300	255	19	74.0	26.0	99.3	0.7
2817	3:19:38	299450	281	18	85.3	14.7	96.7	3.3
2818	6:00:14	540350	713	14	74.3	25.7	77.7	22.4
2904	2:45:09	247725	389	30	90.4	9.6	96.0	4.0
2905	3:15:36	293400	325	20	75.9	24.1	95.1	4.9
2907	2:20:20	210500	257	15	79.3	20.7	94.6	5.4
2908	3:53:21	350025	503	24	74.6	25.4	97.7	2.4
2918	1:21:23	122075	143	7	87.3	12.7	86.7	13.3
2940	3:18:10	297250	402	17	77.8	22.2	88.5	11.5
2959	2:25:17	217925	265	24	64.1	35.9	87.9	12.1
2960	3:31:54	317850	340	22	82.3	17.7	96.2	3.8
2977	2:45:03	247575	447	32	82.7	17.3	94.8	5.2
2978	3:35:27	323175	371	20	78.2	21.8	96.3	3.7
2992	2:08:36	192900	149	2	100.0	0.0	100.0	0.0
2995	2:56:59	265475	580	9	91.2	8.8	71.2	28.8
3011	2:01:42	182550	315	25	62.4	37.6	87.8	12.2
3012	3:37:10	325750	365	24	74.7	25.3	90.1	9.9
3013	4:15:16	382900	501	19	82.1	17.9	80.2	19.8
3014	2:47:22	251050	270	20	80.0	20.0	92.1	7.9
3015	3:02:44	274100	252	13	80.1	19.9	93.8	6.2
3017	3:05:36	278400	291	19	77.3	22.7	69.0	31.0
3020	3:41:58	332950	390	14	86.7	13.3	73.0	27.0
Mean	**3:05:07**	**277672**	**376**	**19**	**80.1**	**19.9**	**89.9**	**10.1**

Table 1 relates the features of the video dataset used in the experimental evaluation with the achieved results with the given video. For each one, *video* column denotes the video identification label that corresponds to a plenary meeting, *length* represents duration time expressed in hours, minutes and seconds, *frames* indicates its number of frames, *shot* shows the number of shot cuts automatically detected (see section 3.2.1), *speakers* denotes the annotated number of different speakers who intervened, and *TRR*, *FRR*, *TDR* and *FDR* correspond to the four rates defined above.

The results show that for the experimental dataset, the mean percentage values of TRR and TDR are respectively 80.1% and 89.9%. These results are promising, in fact the system performs rather well taking into account the large number of different shots (a mean value of 376 shots per video), diverse speakers (a mean value of 19 speakers) who intervene in the different videos and considering that they are frequently looking down reading the speech in most shots hindering face detection, and the playback time of each video (a mean value of three hours). For two specific videos, 2771 and 2992, the TRR is 100%. Although the fact that these results include one video with the lowest number of speakers, two, it is worth noting that the different pose variations of all speakers are correctly detected throughout the different 149 shot cuts for over two hours of video recording. Considering other videos where low values are obtained for the TRR such as video 2791, 2799, 2959 and 3011, it is observed that they all have more than 23 different male speakers and most of them use similar clothes.

Fig. 10 Example of two true re-identifications and one incorrect re-identification

Comparing with the results reported in the re-identification literature, it can be appreciated that our TRR values are not so high in specific videos used in our experimental dataset; however, there is an important difference with those works. In the re-identification literature, the results are obtained on standard datasets as VIPeR [22], CAVIAR4REID [12] or i-LIDS [56]. Those datasets provide the probe and gallery sets with only the cropped image of the person of interest after the background has been removed. For instance, in [22] the re-identification process is performed in a video scenario that contains only one person per video, whilst the approach proposed in this paper deals with real unrestricted scenario where the person to be re-identified can be in the seat surrounded by many other deputies (see Fig. 9) which may act as distractor elements for the person detection stage. At the same time, the misalignment in the detection stage is a source of errors that are not present in re-identification datasets where persons are perfectly cropped. These features affect negatively to the TRR, decreasing its value, but if we focus on the TDR, it can be observed that the high value achieved 89.1% implies the speakers are well distinguished. Considering both TRR and TDR values, it can be inferred that the behavior of the system is able to distinguish among speakers (high TDR) but it can assign more than one *id* to the same speaker (low TRR) in certain situations. However, this situation is easily overcome with the aid of a human operator.

Fig. 10 shows an example of the matching process for four different shots. The first row corresponds to the shot of the first appearance of a specific speaker. In the second and third rows, correct re-identification of the speaker in subsequent shots are shown, while in the last row there is an example of incorrect re-identification because this shot was given the same *id* label by the system than the other shots presented in the figure.
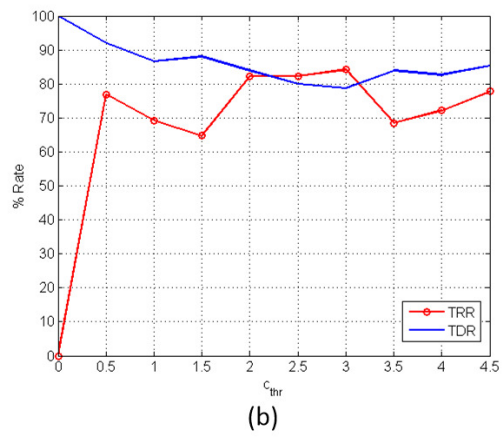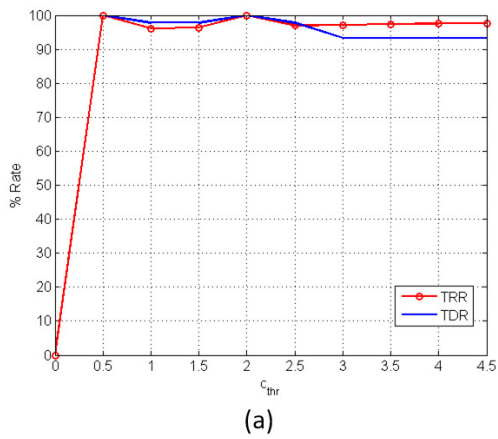
Fig. 11 Variation of TRR and TDR according to the settings of different values of clothes threshold (c_{thr}):
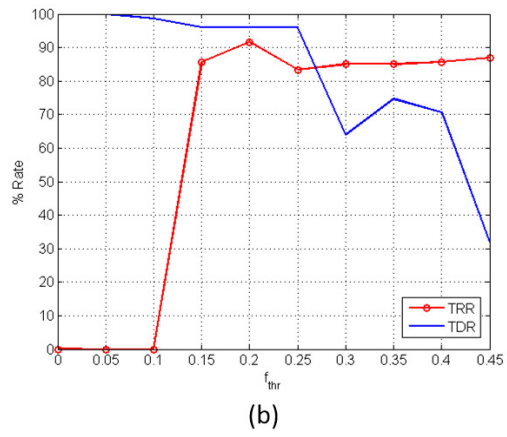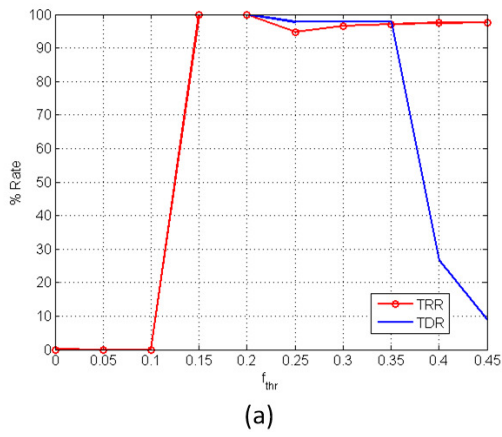**(a) video 2790, (b) video 2918**

Fig. 12 Variation of TRR and TDR according to the settings of different values of face threshold (f_{thr}):
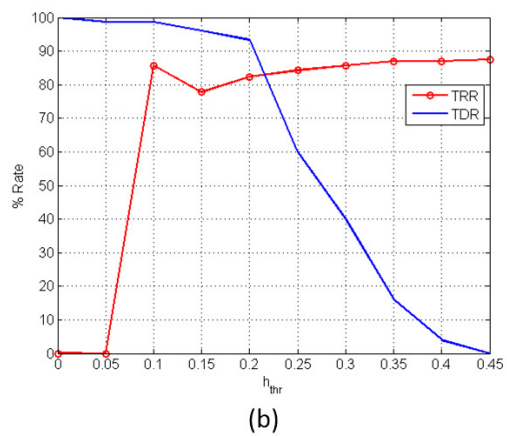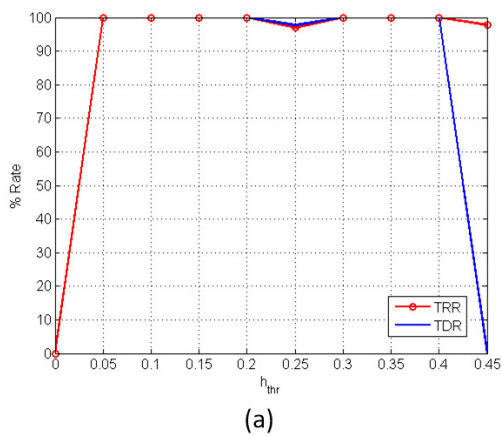**(a) video 2790, (b) video 2918**

Fig. 13 Variation of TRR and TDR according to the settings of different values of hair threshold (h_{thr}):
(a) video 2790, (b) video 2918

Finally, we have tested the influence of the thresholds values setting in the re-identification performance. The re-identification process depends on three decision thresholds (eq. 9) that determine the addition of new identities based on clothes, face and hair features. Fig. 11 shows the value of TRR and TDR for videos 2790 and 2918 for different values of the clothes threshold, c_{thr}. In order to avoid the influence of the other two thresholds, they have been set up to a high value, f_{thr} = 10000 and h_{thr} = 10000. The behavior is similar for both videos, TRR starts at 0% and exhibits an increasing tendency as the threshold increases while TDR presents the opposite behavior, starting at 100% and decreases as the threshold increases its value. Fig. 12 and Fig. 13 present the values of TRR and TDR for the same videos for different values of the face and hair thresholds, f_{thr} and h_{thr}. As previously, the effect of the other two thresholds have been disabled assigning a high value (10000) to them. The behavior for TRR and TDR is the same that the one observed for the clothes threshold. The variation of TRR and TDR for the three thresholds is as expected. For low values of the thresholds a large number of identities is created according to (eq. 9), so high distinction rates are obtained because the system assigns different identities to different speakers. On the other hand, low re-identification rates are achieved because the same speaker is given several identities. For high values of the thresholds a low number of identities are created according to (eq. 9) and thus lower distinction rates and higher re-identification rates are obtained.

5 Conclusions and future work

In this paper, we have described the main functionalities and architecture of a multimedia system to produce and deliver video fragments on demand linked with complete parliamentary information and public feedback that enhances traditional live streaming video solutions on parliamentary websites. A prototype system has been developed for the Canary Islands Parliament (Spain). This system has been designed and validated with parliamentary specialists; therefore, the results are applicable to other parliamentary and government institutions around the world. Feedback and validation data have been retrieved from regular interviews with the CIO of the Canary Islands Parliament to identify the usefulness and benefits of the new approach. The main conclusions of the CIO reflect the importance of the results achieved on the website interface and the improvement of the performance for the parliamentary experts:

- *Website interface:* the CIO of the Canary Islands Parliament affirms that the new interface proposed for the parliamentary website to search for specific information on plenary meetings has been shown to be easy and usable, enabling a straightforward visualization of the results: background information, video fragments on demand, keywords, geolocalization of public opinion, and summaries on overall opinion polarity. This new approach enhances traditional streaming systems in terms of providing the specific offset of videos and detailed and dynamic information, which results in a greater level of transparency and accountability. Providing dynamic information helps users to understand the meaning of the legislative initiative being debated with the description of the status of the related parliamentary documentation to the intervention at all times (procedural actions undertaken to date, related initiatives, the decision-making procedure adopted, and votes on the legislative initiative being debated). Likewise, the speaker information (legislative mandate, geographical constituency, membership to parliamentary body and committees, role played in the parliament

and legislative initiatives proposed) is also considered helpful while the speech fragment is being heard. Furthermore, it is considered useful to be able to visualize Eurovoc descriptors, which supply accurate descriptors to classify speakers' speeches. The visualization of twitter posts related to the speaker's speech and the geolocalization of these posts provide meaningful feedback of citizens on specific speeches.

- *Performance*: the CIO of the Canary Islands Parliament highlights the importance of the automation of the computation of reference speakers to be annotated in the annotation subsystem by the parliamentary experts. The proposed shot detection approach based on the KL divergence of color distribution between consecutive shots has shown itself to be robust to gentle changes in field of view and zoom of the cameras. The re-identification of speakers has been based on discriminate features like clothes color, face and hair. The experimental evaluation of the re-identification approach has been focused on four different rates. The results obtained show that the proposed approach is robust enough for the purpose of determining the different speakers who intervene in plenary meetings. As a consequence, the working time is reduced considerably for parliamentary experts as they only need to process the key representative frames to be annotated and not the whole videos.

We continue to work towards completing the multimedia system in different aspects, which will form our future lines of research with a fourfold purpose. First, enrich the current multimedia system for users with new political data in order to compare politicians in terms of political accountability based on benchmark indicators. Second, enhance the current multimedia system exploiting data analysis on new public-related opinion sources and media articles related to specific plenary sessions. Third, the use of other discriminatory features based on soft biometrics, like the gender of the speaker, which will allow us to make an initial split in the identification process of determining the different speakers who intervene in plenary sessions. Fourth, the incorporation of speech features to complement and improve the segmentation tool.

Acknowledgements

References

1. Abdallah, S., & Raimond Y. (2007). The Timeline Ontology. http://purl.org/NET/c4dm/timeline.owl#. Accessed May 2015.
2. Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, nº 12, pp. 2037- 2041.
3. Barra-Chicote R, Pardo J, Ferreiros J. & Montero J (2011) Speaker diarization based on intensity channel contribution. In IEEE Transactions on Audio, Speech & Language Processing, 19(4): 754-761
4. Berners-Lee T, Hendler J, Lassila O (2001) The semantic web. The Scientific American Magazine, 284(5):28-37.
5. Berners-Lee T (2006) Linked Data. International Journal on Semantic Web and Information Systems, vol 4, nº 2:1-72.
6. Bertot J, Jaeger P, Grimes M (2010) Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. Government Information Quartely, 27(2010), 264-271.

7. Bixtext API. (2014) http://www.bitext.com/. Accessed May 2015.

8. Bizer C, Heath T, Berners-Lee (2009) Linked Data – The story so far. International Journal on Semantic Web and Informatyion Systemes, 5(3):1-22.

9. Breslin G, Harth A, Bojars U, Decker S (2005) Towards semantically-interlinked online communities. In 2nd European Semantic Web Conference (ESWC 2005), pp. 500-514.

10. Brickley D, Miller L (2014) FOAF Vocabulary Specification 0.99. Namespace Document 14 January 2014 – Paddington Edition, FOAF Project. http://xmlns.com/foaf/spec/. Accessed May 2015.

11. Bulterman D, Grassel G, Jansen J, Koivisto A, Layaida N, Michel T, Mullender S, and Zucker D (ed) (2005) Synchronized multimedia integration language (SMIL 2.1). W3C recommendation. World Wide Web.

12. Cheng D, Cristani M, Stoppa M, Bazzani L, Murino V (2011) Custom pictorial structures for re-identification. In Brithish Machine Vision Conference, Dundee.

13. Castrillón Santana M, Déniz O, Hernández, D, Lorenzo J.A (2011) Comparison of Face and Facial Feature Detectors based on the Viola-Jones General Object Detection Framework. Machine Vision and Applications, vol. 22, issue 3.

14. Cong Truong D, Khoudour L, Achard C, Meurie C, Lezoray O (2010) People re-identification by spectral classification of silhouettes. Signal Processing, vol.90, pp.2362-2374, doi:10.1016/j.sigpro.2009.09.005

15. Cover M, Thomas A (2006) Elements of Information Theory. Wiley Interscience.

16. DBpedia (2014) The DBpedia Ontology 3.9. http://wiki.dbpedia.org/Ontology. Accessed May 2015.

17. Déniz O, Bueno G, Salido, J, Torre D (2011) Face recognition using histograms of oriented gradients. Pattern Recognition Letters, vol. 32, nº 12, pp. 1598-1603.

18. Dublin Core Metadata Innovation (2006). DCMI term declarations represented in RDF schema language.

19. EBU/ETSI (2008) ETSI TS 102 822-3-1: TV-anytime. http://www.tv-anytime.org/. Accessed May 2015.

20. Eurovoc (2014). EU's multilingual thesaurus. http://eurovoc.europa.eu/drupal/. Accessed May 2015.

21. Global Centre for ICT in Parliament United Nations Department of Economic and Social Affairs (2012) World e-Parliament Report 2012. In: Inter-Parliamentary Union Press, by the Division for Public Administration and Development Management of the United Nations.

22. Gray D, Tao H (2008) Viewpoint Invariant Pedestrian Recognition with Ensemble of Localized Features. Proc. European Conference on Computer Vision (ECCV 2008).

23. Hausenblas M, Troncy R, Burguer T, Raimond Y (2009) Interlinking Multimedia: How to Apply Linked Data Principles to Multimedia Fragments. In: LDOW2009, Madrid, Spain.

24. Hausenblas M, Karnstedt, M (2010) Understanding Linked Open Data as a Web-Scale Database. In Proc. Second Int'l Conf. Advances in Databases Knowledge and Data Applications (DBKDA), pp. 56-61.

25. Inter- Parliamentary Union (IPU) (2009) Guidelines for the content and structure of parliamentary web sites. In: Inter-Parliamentary Union Press, United Nations.

26. ISO/IEC (2002). Information technology- Multimedia Content Description Interface. ISO/IEC 15938-1:2002.

27. ISO/IEC (2006) Information technology-MPEG-21, part17: fragment identification of MPEG resources. ISO/IEC 21000-17:2006.

28. Knox M, Friendland G (2010) Using oriented optical flow histograms for multimodal speaker diarization. In Proceedings of ISCA Interspeech, Makuhari, Japan, September 2010, pp. 290-293.

29. Kontopoulos E, Berberidis C, Dergiades T, Bassiliades N (2013) Ontology-based sentiment analysis of twitter posts. In Journal of Expert Systems and Applications, http://dx.doi/org/10.1016/j.eswa.2013.01.001.

30. Kumar N, Berg A, Belhumeur P, Nayar S (2011) Describable visual attributes for face verification and image search. In IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 33, Issue 10, pp: 1962-1977.

31. Lefevre S, Vincent N (2007) Efficient and robust shot detection. Journal of Real-Time and Image Processing, vol. e, pp. 23-34.

32. Li Y, Wald M, Wills G (2011) Interlinking Multimedia Annotations. ACM Web Science Conference 2011, Koblenz, Germany.

33. Li Y, Wald M, Wills G (2012) Applying linked data in multimedia annotations. International Journal of Semantic Computing, 6(3), 203-237.

34. Lorenzo-Navarro J, Castrillón-Santana M, Freire-Obregón, Ramón-Balmaseda E (2015) Stripe based clothes segmentation, Workshop on Video Analytics for Audience Measurement (VAAM15).

35. Mannens E, Derusen D, Troncy R, Pfeiffer S, Parker C, Lafon Y, Jansen J, Haunsenblas M, Vand de Wall R (2012) A URI-based approach for addressing fragments of media resources on the Web. Multimedia Tools and Applications Journal, Springer-Verlag, 59: 691-715. DOI 10.1007/s11042-010-0683-z.

36. Miró X, Bozonnet S, Evans N, Fredouille C, Friendland G (2012) Speaker diarization: A review of recent research. In IEEE Transactions on Audio, Speech & Language Processing, 20(2):356-370.
37. OpenLink Software (2014). Virtuoso Universal Server. http://virtuoso.openlinksw.com/. Accessed May 2015.
38. Saif, H, He Y, Alani, H (2012) Alleviating data sparsity for twitter sentiment analysis. In 2nd Workshop on making sense of microposts (#MSM2012): Big things come in small packages at World Wide Web (WWW2012), pp. 2-9.
39. Schandl B, Haslhofer B, Burger T, Langegger A, Halb W (2012) Linked data and multimedia: the state of affairs. Multimedia Tools Applications Journal, Springer-Verlag 59:523-556. DOI 10.1007/s11042-011-0762-9.
40. Sikora T (2001) The MPEG-7 visual standard for content description – an overview. IEEE Transactions on Circuits and Systems for Video Technology, 11(6), 696-702.
41. Stankovic M, Passant A, Laublet P (2009) Directing Status Messages to their Audience in Online Communities, MALLOW2009.
42. Studer R, Benjamins R., Fensel, D (1998) Knowledge engineering: Principles and methods. IEEE Transactions on Data and Knowledge Engineering, 25(1-2), 161-197.
43. Tapaswi, M, Bauml M, Stiefellhagen, R (2012) Knock! Knock! Who is it? Probabilistic person identification in tv series. Conference on Computer and Pattern Recognition (CVPR).
44. Twitter (2013) Using the Twitter Search API. https://dev.twitter.com/docs/using-search. Accessed May 2015.
45. Vatant B. & Wick M. (2012). GeoNames Ontology. http://www.geonames.org/ontology/documentation.html. Accessed May 2015.
46. Vezzani R, Baltieri D, Cucchiara R (2013) People reidentification in surveillance and forensics: A survey. ACM Computing Surveys. Vol. 46, issue 2, article nº 29. DOI: 10.1145/2543581.2543596.
47. Vila M, Bardera A, Xu Q (2013) Tsallis entropy-based information measures for shot boundary detection and keyframe selection. Signal, Image and Video Processing, vol. 7, nº 3, pp. 507-520.
48. Viola P, Jones M (2004) Robust real-time face detection. International Journal of Computer Vision, 57(2), 137-154.
49. W3C 2004 Web Ontology Language. http://www.w3.org/TR/owl-features/. Accessed May 2015.
50. W3C (2008) Semantic Web Activity. SPARQL query language for RDF. http://www.w3.org/TR/rdf-sparql-query/. Accessed May 2015.
51. W3C Media Fragments Working Group (2010) Media Fragments working group. http://www.w3.org/2008/WebVideo/Fragments, Accessed May 2015.
52. W3C (2012a) Media Fragments URI 1.0. http://www.w3.org/TR/media-frags/. Accessed May 2015.
53. W3C (2012b) W3C Media Annotation Working Group. Ontology for Media Resources 1.0. http://www.w3.org/TR/mediaont-10/. Accessed May 2015.
54. W3C (2014a) W3C Semantic Web. Activity Resource Description Framework 1.1 (RDF) http://www.w3.org/RDF/. Accessed May 2015.
55. W3C (2014b) W3C Semantic Web activity. RDF Schema 1.1. http://www.w3.org/TR/rdf-schema/. Accessed May 2015.
56. Wang T, Gong S, Zhu X, Wang S (2014) Person Re-Identification by Video Ranking. Proc. European Conference on Computer Vision (ECCV 2014)

Elena Sánchez-Nielsen is Associate Professor of the Department of Informatics and Systems at University of La Laguna. She received her M.Sc. and Ph.D. degrees in Computer Science and Artificial Intelligence from the University of La Laguna in 1999 and 2003 respectively. She is involved in European, national and regional projects related to computer vision, multimedia content, smart information systems, eGovernment and eParticipation. Her research interests include multimedia systems, smart information systems, web semantic technologies, eGovernment, computer vision for human interaction and social media.

Francisco Chávez-Gutiérrez is CIO of the Canary Islands Parliament and responsible for the development of ICT projects in this institution. The last ICT project deployed in this institution has been the production and recording multimedia system to stream plenary meetings and events. He received his B.Sc. degree in Computer Science from the University of Las Palmas de Gran Canaria in 1995 and his M.Sc. in 2007 from the University of La Laguna. Currently, he is developing his PhD in the University of La Laguna. In recent years, he has collaborated with the University of La Laguna in R&D projects related to the adaptation of new technologies in information systems of Public Administration to enhance their functionalities. His research interests include eGovernment, modernization of parliamentary information systems, audiovisual contents and web semantic technologies.

Javier Lorenzo-Navarro is Associate Professor of the Department of Informatics and Systems and Researcher of the Institute SIANI (Intelligent Systems and Numerical Applications in Engineering) at University of Las Palmas de Gran Canaria (ULPG). He received his M.Sc. in Informatics in 1992 from ULPGC. In 2001 he obtained his PhD from ULPGC with the work entitled "Feature Selection in Machine Learning based on Information Theory" for which he obtained the Best PhD Award. His research interests include Computer Vision for Human Interaction, Machine Learning and Data Mining and Wind Farm Power Prediction.

Modesto Castrillón-Santana received the M.Sc. and Ph.D. degrees in Computer Science from Las Palmas de Gran Canaria University (ULPGC), in 1992 and 2003 respectively. His research activities focus particularly on the automatic facial analysis problem, but covering also different topics related to image processing, perceptual interaction, human-machine interaction and computer graphics. Currently, he is an associate professor at ULPGC.