Automatic generation of matter-of-opinion video documentaries

Stefano Bocconi [a,*], Frank Nack [b], Lynda Hardman [b,c]

[a]*Università di Torino, Dipartimento di Informatica, Corso Svizzera 185, 10149 Torino Italy*
[b]*CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands*
[c]*Technical University of Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*

Abstract

In this paper we describe a model for automatically generating video documentaries. This allows viewers to specify the subject and the point of view of the documentary to be generated. The domain is matter-of-opinion documentaries based on interviews. The model combines rhetorical presentation patterns used by documentary makers with a data-driven approach. Rhetorical presentation patterns provide the viewer with an engaging viewing experience, while a data-driven approach can be applied to growing media repositories. To date, the modeling of rhetoric has been achieved in a top-down manner using closed repositories, while data-driven generation approaches were unable to implement non-trivial rhetorical presentation patterns. We describe an implementation of our model in a system, Vox Populi, and apply it to an online documentary shot by a group of independent amateur documentarists.

Key words: video editing, semantic media annotations, Toulmin-based argumentation, documentary rhetoric, user-specified point-of-view

1. Introduction

We are used to viewing a documentary as a fixed static artifact. This artifact is the product of a director who crafted it for us, using footage recorded for the purpose of making a film. This scenario does not include any viewer intervention except at the last stage, when the viewer can decide to view the documentary or not, and even then the choice is pretty limited to watch (a part of) it or ignore it. Why is this a problem? Because documentaries are meant to inform as well as entertain. The video material collected during shooting is quantitatively much more than the material that is selected for the final version. In the case of a documentary, this can mean that large amounts of footage with different themes, topics and arguments will never be seen by the viewer, if only because of time limits. Moreover, when a documentary is about a matter-of-opinion issue, a documentarist has the power to build a strong argument either for or against an opinion by selecting and editing different footage from the available material. A documentarist determines a documentary's content and point of view for all viewers, where the viewers themselves would probably have made other choices. On the other hand, making all footage available is not a suitable alternative, because a viewer is unlikely to be willing to watch hours of video with no story or theme, and most likely no apparent relation between one sequence and the following one. Documentaries still offer good narrative models to show content in a way that does not overwhelm the viewer.

A solution to the above mentioned problems could be represented by an automatic approach towards documentary generation. Such an approach should combine the strengths of traditional documentary making and automatic video generation: the former is capable of presenting issues to the viewer in a way that is in-

* Corresponding author. Tel: +39 011 6706834
 Email addresses: Stefano.Bocconi@di.unito.it
(Stefano Bocconi), Frank.Nack@cwi.nl (Frank Nack),
Lynda.Hardman@cwi.nl (Lynda Hardman).

formative and interesting at the same time, while the latter allows the documentarist to provide viewers with documentaries dynamically generated according to their interests. Automatic video generation not only has advantages for the viewer, but also for the documentarist. A video generation system could help the documentarist by automatically presenting the material, freeing her from the need to select and edit the footage, which is a difficult and time-consuming task. Different documentaries can then be generated from the same footage, facilitating reuse of the media assets, and allowing new footage to be added at a later stage. Automatic video generation can transform a documentary from a static final product into an evolving up-to-date video document.

To date, there is no single approach capable of combining the advantages of human authoring with the potential benefits of automatic video generation, which we summarize in the following three points. While there are systems that satisfy one or more points, there is no single approach that includes all of them:

(i) generate documentaries on matter-of-opinion issues which use presentation forms as a documentarist would do (as, for example, Splicer [20] does).

(ii) allow the viewer to select both the content of the generated documentary and the documentary's point of view (as, for example, Terminal Time [14] does).

(iii) allow the documentarist to collect material to be used for documentaries, without having to specify how this material should be presented to the viewer, and to add footage at a later stage (as, for example, ConTour [16] does).

In this paper we focus on documentaries about matter-of-opinion issues, where opinions are mainly expressed by people being interviewed. In this type of documentary the drawbacks of having a final static version are evident: especially when the number of interviewees is high, for time constraints some of the interviewees' answers will not be selected, and possibly some opinions will never be displayed to the viewer. We propose an automatic video generation approach that allows the viewer to potentially see all material shot for a particular documentary, not only what a documentarist would have selected. The content of such a documentary is determined by the viewer choosing a particular subject she is interested in and a point of view. The material is then organized according to presentation forms also used by documentarists. Automatic video generation allows the repository containing the raw footage to grow by adding relevant material, and both new and old ma-

terial to be used to generate new documentaries.

The paper is organized as follows: in section 2 we describe the domain of our research, namely matter-of-opinion documentaries based on interviews. In section 3 we examine related work to gather information on the type of annotations we need for our approach. In section 4 we describe our video generation model and give an example of a generated documentary. In section 5 we provide some conclusions and some directions for future work.

2. What is a documentary?

We base our view on documentaries on the definition of the documentary genre from Bordwell and Thompson ([6], p. 128):

"*A documentary film purports to present factual information about the world outside the film.*"

This definition stresses that the most distinctive feature of a documentary is the intention of presenting informative content. Typically, a film labeled as *documentary* leads the viewer to assume that the persons, places and events exist and that the information presented is trustworthy ([6], p. 128). Rabiger ([19], pp. 3-4) says that documentaries explore actual people and actual situations and they always reflect a profound fascination with, and respect for, actuality.

If we look at how documentaries present this factual information, Bordwell distinguishes three types of form: the **narrative form**, the **categorical form** and the **rhetorical form** ([6] p. 132). Documentaries often use more than one of these forms, but usually one type is predominant over the others. The rhetorical form is particularly relevant for the type of content we are interested in, i.e. matter-of-opinion issues. In using the **rhetorical form**, a documentary aims at persuading the audience to adopt an opinion about the subject, usually a matter-of-opinion issue. In a rhetorical documentary, the documentarist tries to make her **point of view** seem the most plausible by presenting different types of arguments and evidence ([6], p. 140).

Rabiger ([19] pp. 8-9) discusses three different ways a documentarist can behave, depending on her respect for the audience: at the lowest level of respect there is the **propagandist**, who wants to condition the audience, showing only the evidence supporting predetermined conclusions. Moving up the scale of respect, there is the **binary communicator**, who gives "equal coverage to both sides" in any controversy. Rabiger says that this form considers the audience as a passive mass to be informed and entertained, but not challenged to make

judgments. At a higher level, is the **mind-opener**, who aims not at conditioning or diverting but at expanding the viewer's mind, by presenting something in all its complexity, never patronizing or manipulating either the subjects or the audience.

Presenting opposing **positions** has a number of advantages for the documentarist. It can be used as a technique to make the audience want to see what will happen next. The purpose of this is to introduce some level of dramatic conflict into the structure of the documentary. Dramatic conflict is a structural tension that keeps the outcome of the documentary somewhat in doubt and keeps the audience interested ([11] pp. 298-299).

Typically, when people state their opinion or position, they do so with a discourse intended to persuade or prove that their conclusions are correct. We call such a discourse an **argument**. The Greek philosopher Aristotle (in his book "Rhetoric") classified means of persuasion into logos, pathos and ethos. These means of persuasion are used by a *speaker* (or author) who tries to convince an *audience*. In the following we will concentrate on **logos**, which appeals to logic or reason. A logos argument is based on factual data and on the conclusions that can be drawn from it. The audience should accept the argument because it sounds rational. In documentaries, the audience is the collection of viewers. The speaker can be the documentarist herself (when she appears or speaks in the documentary), the narrator, if there is one, or, in the case of interviews, the interviewees. Interview documentaries (also called "talking heads" documentaries) record testimonies about events or social movements. When the subject of the interviews is controversial, the way arguments are presented is particularly important for the audience to decide whom to believe, because evidence to determine the truth might be lacking ([11] p. 63).

3. Annotations for video generation

After having defined our domain of application, in this section we investigate annotations, since the characteristics of an automatic generation system depend to a large extent on the type of annotation *structure* the system uses. Automatic video generation systems use descriptions (annotations) of the media items in order to make decisions about how to create a video sequence. The structure of annotations is composed out of two parts:
– the structure of the **description** (e.g. a film can be described by fields, such as *title*, *director*),
– the structure of the **values** used to fill the description

(e.g. "A clockwork orange" can be the value of the field *title*).

We introduce three different types of *description structures* (keyword, property and relation based) and four different types of *value structures* (free text, taxonomies, thesauri and ontologies) to represent the range of possible annotation structures.

In structure based on *keywords* (**K-annotations**), each item is associated with a list of terms (words) that represent the item's content. The association to the content is unspecified: for example, an annotation consisting of two keywords "*Rembrandt*, *painting*" can indicate that the annotated item represents a painting made by Rembrandt or a painting about Rembrandt. In this sense, *K-annotations* are ambiguous.

In structures based on *properties* (**P-annotations**), items are annotated with property-value pairs, e.g. *subject-NightWatch*, *creator-Rembrandt*, *date-1642*. Categories allow the disambiguation of cases such as the one above: using *P-annotations*, *Rembrandt* would be either the value of the property *creator* or of the property *subject*.

In structure based on *relations* (**R-annotations**), items are annotated with property-value pairs as in *P-annotations*, only that some of these values are references to other annotations, e.g. [item *X represents Rembrandt*] *hasOffspring* [item *Y represents Titus*] indicates that Titus (represented by media item *Y*) is son of Rembrandt (represented by media item *X*).

When annotating an item, keywords, properties and relations need to be assigned a value. In the case of *R-annotations*, values are a reference to other annotations, while for *K-annotations* and *P-annotations*, values can be chosen from four different types of sources ([1], [21]):

(i) **free text** gives the annotator complete freedom to choose the word that better expresses the content. Terms have no relation to each other.

(ii) a **taxonomy** consists of terms and their hierarchical structure. Each term in a taxonomy is in one or more parent-child relationships to other terms in the taxonomy. A taxonomy does not provide associational relationships between the terms.

(iii) a **thesaurus** consists of terms and their relationships. Relationships within a thesaurus can be hierarchical (as in a taxonomy) and associational (e.g., term A *is related to* term B).

(iv) an **ontology** consists of concepts, which have hierarchical and associational relationships, as in a thesaurus. An ontology attempts to define concepts and show the relationships between concepts, whereas a thesaurus attempts to show the

relationships between terms ([1], p. 8). Unlike the terms in a thesaurus, concepts in an ontology can have properties and formal constraints on how they can be used together.

We call taxonomies, thesauri and ontologies **controlled vocabularies** because values have some constraints due to the structure they have, while free text has none. In the next sections we examine different structures of annotations by discussing document generation systems from literature that use those structures.

3.1. K-annotations

CONTOUR [16] was developed to support evolving documentaries, i.e. documentaries that could incorporate new media items as soon as they were made. The underlying philosophy was that some stories keep evolving, and so should the documentaries describing them. The system was used to support an evolving documentary about an urban project in Boston [1].

ConTour has a twofold aim: for the author, to provide a framework for gathering content and making it available without having to specify explicitly how (and in what order) the user should view the material; for the user, to support visual navigation of the content.

ConTour allows the author to create and expand the repository by adding material to it. The author is required to attach keywords (called *descriptors*) to each media item. The goal of the descriptors is to capture abstract ideas or elements relevant for the documentary story, e.g. names of people or places. Descriptors are created beforehand, with values belonging to the categories of character, time, location and theme. Referring to our classification, ConTour's *value structure* can be considered a simple version of a taxonomy that has four top classes to which all values belong.

Keywords in ConTour relieve authors from the process of defining explicit relationships or links between units of content. Instead, the author connects media items only to keywords. By doing so, the author defines a potential connection between a media item and other media items that share that keyword. Since there are no explicit links between the clips, sequencing decisions are made during viewing, based on the implicit connections via the keywords. Deferring sequencing decisions in this way has as a consequence that the base of content is extensible. Every new media item is simply described by keywords, rather than hardwired to every other relevant media item in the system. In this way, the potentially exponentially-complex task of adding content is managed and requires a constant effort. In this sense ConTour's approach is data-driven, and the key factor is that links are created automatically by the system, and not by the author.

On the other hand, by using keywords ConTour can only determine to what degree two media items are related: from unrelated (if they have no descriptors in common) to totally related (if they have the same descriptors). This relation cannot be further specified by the system: is one media item providing further information with respect to another, or is it contradicting the information presented by the other? This limitation is inherent to keywords, as Cleary and Bareiss also demonstrated ([7], p. 35). Other systems that use *K-annotations* are Lev Manovich's SOFT CINEMA [2] and the KORSAKOW SYSTEM [3], systems that edit movies in real time by selecting media items from a database.

3.2. P-annotations

SEMINF [13] is a system that creates presentations using media items from annotated repositories. SemInf uses repositories belonging to the Open Archives Initiative [12], which are annotated with the Dublin Core (DC) schema [9]. This schema is designed to be very simple in order to facilitate a widespread adoption, with little overhead in annotating the material. It was conceived by and for librarians, and it contains properties for classifying items in a library, e.g. *creator*, *date*, *description*, *format*, *title*.

SemInf's main concern is how to layout media items so that the viewer understands the semantics of the presentation. To determine this, SemInf infers relations between the media items it has to display. For example, if media item *X* is annotated with the property-value pair *DC.creator - Rembrandt* and media item *Y* is annotated with *DC.description - Rembrandt*, SemInf infers that *Y* represents the *creator* of *X*, or, in other words, the relation between the two media items is (person depicted in) *Y creates X*. In this way, a set of relations are inferred of the type *X creates Y*, *X describes Y*, *X colleagueOf Y*, etc. These relations are then translated into spatial/temporal relations in the presentation, driving the layout of the items on the screen. For example, *X creates Y* is translated to *X spatialLeft Y*, causing *X* to be displayed on the left of *Y*.

---

[1] The "Big Dig" project was aimed at relieving Boston, MA, from a huge traffic problem caused by an elevated six-lane highway, called the Central Artery, that ran through the center of downtown.

[2] http://www.softcinema.net/
[3] http://www.korsakow.com/ksy/index.html

SemInf shows that *P-annotations* make the process of inferring relations between annotated items possible, although the relations SemInf is able to infer are very simple. This is due to DC's simplicity, since DC was designed to find items in a (digital) library, more than to support presentations about those items. Furthermore, DC annotations use values from free text. Using free text different words can mean the same, as well as the same word can mean different things. This hinders the inferencing: for example, to determine that *X created Y*, SemInf checks whether the condition *Y.creator == X.subject* holds. Ambiguity in the values causes relations to be created incorrectly.

From examining SemInf, we can draw two conclusions. The first is that *P-annotations* allow inferencing relations. The second is that inferencing needs annotations that use a controlled vocabulary of values, such as a taxonomy, a thesaurus or an ontology.

3.3. R-annotations

DISC [10] is a multimedia presentation generation system for the domain of cultural heritage. This system uses the annotated multimedia repository of the Rijksmuseum [4], to create multimedia presentations. Disc uses *R-annotations* of the form [media item *X represents Rembrandt*] *hasOffspring* [media item *Y represents Titus*], [media item *X represents Rembrandt*] *hasTeacher* [media item *Z represents PeterLastman*]. The system uses the stereotypical structure of well-established narrative genres, such as biography, to organize the content for a presentation. The content is selected by rules. When an annotation fulfills a rule, the corresponding item is included in the presentation. For example, a biography about an artist typically discusses the artist's teacher, if there was one. To generate a biography about Rembrandt, Disc executes a "teacher" rule that verifies whether the item annotated as representing Rembrandt also has a relation *hasTeacher*, i.e. *Rembrandt hasTeacher z*, where *z* is represented in media item *Z*. If this is the case, as it is in Rembrandt's case with Pieter Lastman, the media item *Z* representing Rembrandt's teacher is included in the presentation, in the section talking about Rembrandt's career.

Repositories annotated with *R-annotations* can be represented as a graph whose nodes reference the items and whose edges are the relations between them. We call this graph a *Semantic Graph*. The *Semantic Graph* can then be traversed to generate presentations com-

posed of the annotated items. In Disc's case, the semantic graph is traversed by rules to generate narratives. The semantic graph provides the story space within which the rules select narrative presentations. The drawback of manually creating a semantic graph is that all possible combinations of two items in the repository must be examined to check whether the concepts they represent should be related, as required by approaches using *R-annotations* (see e.g. [22] and the discussion in section 3.1). This fact makes the process of annotating, when the repository is created and each time a new element is added, particularly cumbersome.

3.4. *Conclusions*

As Disc shows, a *Semantic Graph* can be traversed using rules to generate narrative presentation structures. A semantic graph can be given, as in the case of *R-annotations*, or it must be inferred when using other types of annotations. *R-annotations* cannot be used for repositories that can grow with new elements, because of the need to examine all existing items in order to asses whether they should be related to the new one. This goes against point iii in section 1. *K-annotations*, used in ConTour, do not require the documentarist to specify how the items should be related. On the other hand, *K-annotations* only allow the creation of generic associational links between items. To implement a rhetorical form for documentaries (described in section 2), we need to infer the rhetorical relation between two related interview clips, i.e. whether a statement expressed in one interview *supports* or *contradicts* a statement in another interview. *P-annotations* provide a means for describing content and are able to support inferencing the relations between media items, as long as the values are not ambiguous, as in SemInf's case. In our case we can therefore use *P-annotations*, using a *controlled vocabulary*, i.e. a taxonomy, a thesaurus or an ontology. For our approach we use a thesaurus, as motivated in section 4.1.2.

4. A video generation model

We determined in the previous section that we need to use *P-annotations* and a controlled vocabulary to annotate video content in order to capture the information needed for our automatic video generation approach.

The goal of this section is to define which content contained in video must be annotated, and how it must be annotated to implement the rhetorical form described in section 2. The elements of this form are *points of*

view, *positions* and *logos arguments* [5] . Content in video is conveyed by the video track and in the audio track, which are processed by the visual and auditory perception channels, respectively. Metz identifies six media types in video ([23]):

– **visual channel** (video track): **image**, **video** and **writing** (credits, intertitles, subtitles, written materials in a shot)
– **auditory channel** (audio track): **noise**, **music** and **speech**.

These media types can contain two types of information: **Verbal** (information conveyed by language, e.g. speech, writing) and **Non Verbal** (all other information, e.g. noise, music, video).

In order to model the elements of the rhetorical form, we need to determine which media types contain the content relevant for arguments and positions. Having determined where the relevant information is, we have to define how to model it.

We first model arguments based on logos. We then show how to generate a documentary in section 4.2.

4.1. *Modeling logos*

The **logos** technique appeals to logic or reason. Arguments using logos are based on factual data and on the conclusions that can be drawn from it. These conclusions should be accepted by an audience because they sound rational. Logic and rationality require a certain degree of abstraction, and are expressed using language, which can be of any type, e.g. natural language or symbolic language. For these reasons we model logos by modeling *verbal* information. Verbal information is present in speech in the auditory channel and writing in the visual channel, therefore in order to model logos we need to look at speech and writing.

4.1.1. *Statements*

In interview documentaries, most of the verbal information is conveyed by speech, i.e. by the interviewees' answers to questions. A **statement** is a short sentence that captures the sense of what the speaker says, such as "War is not effective", or "Diplomacy cannot be used". A statement can summarize the actual words used by the interviewee while expressing her position. For example, the transcript "I am never a fan of military actions, in the big picture I do not think they are ever a

good thing" can be summarized by the statement "*Military actions are not effective*" or "*Military actions are not good*". Statements can also encode visual and non-visual information which is non-verbal in nature, but can be associated with a verbal message. For example, a video sequence of a river being polluted by a factory can express the statement "*Factories pollute the environment*", although the association is not as strong as for verbal information. Statements do not capture all the semantics contained in the original sentences. This is not necessarily a limitation for our purposes, since we only need to encode sufficient information to represent how arguments can be built, analogous to the approaches adopted by ScholOnto [22] and Splicer [20].

We model statements using a three-part structure: a **subject**, a **modifier** and a **predicate**. The subject (s) represents the subject of the statement, the predicate (p) qualifies the subject and the modifier (m) modifies the relation between the subject and the predicate. A statement is not required to have a modifier (*no mod*), whereas the subject and the predicate are required. The statement "*They are using two billion dollar bombs on ten dollar tents*", for example, is encoded as s:bombing m:not p:effective.

The choice of a three-part structure results from a trade-off between expressiveness (how well a statement represents what is actually said) and computational complexity (how processor-intensive inferencing on these statements is). Using more than three parts would increase the expressiveness but also the computational complexity. In [3] we tried a four-part structure, but we found that with three parts we can describe the clip content with a degree of detail sufficient to represent arguments. AAB [8], which generates abstracts of video programs, such as symphonies or football matches, by selecting salient scenes, also uses a three-part structure.

4.1.2. *Thesaurus*

As discussed in section 3, the terms used for the subject, predicate and modifier must belong to a controlled vocabulary, i.e. the *value structure* must be either a taxonomy, a thesaurus or an ontology. Using a controlled vocabulary allows inferencing of relations between the statements and the corresponding video clips, since the relation between two terms can be used to infer the relation between two statements that contain these terms. Two conflicting interests are at stake in choosing the *value structure*: inferencing is facilitated by constrained structures, such as an ontology, while an annotator's effort is reduced by having a loose structure, such as free text. The ideal compromise is when annotating requires

---

[5] Pathos and ethos arguments need also to be modeled. For space constraints we limit the discussion in this paper to logos arguments, and we refer to [4] for the other two types.

	bomb.	war	peace	dipl.	mil. act.	econ. aid
bombing	Id	Gen.				Opp.
war	Spec.	Id	Opp.	Opp.	Sim.	Opp.
peace		Opp.	Id			
diplomacy		Opp.		Id	Opp.	
military-actions		Sim.		Opp.	Id	
economic-aid	Opp.	Opp.				Id

Table 1

Example of terms and relations between terms contained in the thesaurus for the subject part of the statement.

	no mod	not	never	possibly	once
no mod	Id	Opposite	Opposite		
not	Opposite	Id	Similar		
never	Opposite	Similar	Id		Opposite
possibly				Id	Similar
once			Opposite	Similar	Id

Table 2

Example of terms and relations between terms contained in the thesaurus for the modifier part of the statement. The *no mod* has a positive meaning, being opposite to *not* and *never*.

	effective	waste	useless
effective	Id	Opposite	
waste	Opposite	Id	Similar
useless		Similar	Id

Table 3

Example of terms and relations between terms contained in the thesaurus for the predicate part of the statement.
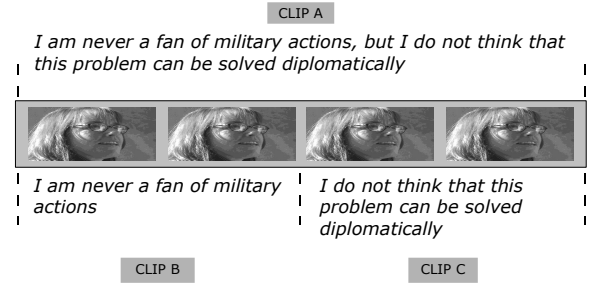

Figure 1. Clip A contains the complete interviewee's answer containing two statements, while Clip B and Clip C segment the answer in two parts of one statement each

the least effort while still supporting the inferencing process. The choice of the controlled vocabulary is therefore dependent on the inferencing mechanism. We need a controlled vocabulary to have relations between terms, because such relations support the process of inferring argumentation relations between the statements, as we show in section 4.2. A taxonomy provides only a hierarchy, but no relations between terms, and it is therefore not suited for our purpose. An ontology can provide relations between terms, as well as properties and formal constraints on how these terms can be used together. Since we only need the relations between terms, an ontology would require an unnecessary modeling effort. In our approach we use therefore a thesaurus. Based on [25], we use the following relations: **Generalization** (Broader Term, hypernym, inverse of *Specialization*), **Specialization** (Narrower Term, hyponym, inverse of *Generalization*), **Similar** (Related Term, holds between synonyms and near-synonyms, symmetric), and **Opposite** (antonym, between two different words of opposite meaning, symmetric) [6].

When annotating the content of a clip with a statement, an annotator is required to use terms belonging to the thesaurus as values for each of the three parts of a statement. The thesaurus can be built while annotating

media items, by inserting the terms used to compose the statements. Existing thesauri, such as Wordnet [15], can provide an aid or a starting point. An annotator also needs to relate each term she uses to the other terms in the thesaurus, using the four thesaurus relations. Because subject, modifier and predicate each play a different role in the statement, terms used for one part of the statement are not related to terms used for another part of the statement. The thesaurus can thus be considered as composed of three independent thesauri, one for each part of the statement. An example of each thesaurus is represented in Tables 1, 2 and 3.

4.1.3. *Clips granularity*

In the case of interview documentaries, the length of the clips that needs to be annotated must be determined as a trade-off between how easy it is to reuse the clip and how representative the clip is to the interviewee's intentions. Longer clips are more self-contained because they establish more context, but for the same reason more difficult to reuse in another context. Longer clips are thus more difficult to use for building different arguments. A finer granularity allows more options for building arguments but risks misrepresenting the interviewee's position. We discuss this issue with an example (see Fig. 1): consider a video interview stating the following: "I am never a fan of military actions, but I do not think that this problem can be solved diplomatically". If this video interview is annotated as a single

---

[6] To make notations easier to display, we also introduce an identity relation **Id** between each term and itself, symmetric.

clip (clip A), it can be used in an argument *for* military actions. If, instead, it is also segmented into the following two clips (clip B): "*I am never a fan of military actions*" and (clip C) "*I do not think that this problem can be solved diplomatically*", clip B can be used in an argument *against* military actions, while clip C can be used in an argument *for* military actions. Using terms from the thesauri shown in Tables 1, 2 and 3 clip B can be annotated with the statement s:military actions m:not p:effective and clip C with s:diplomacy m:not p:effective.

A finer granularity offers thus more options to build arguments. On the other hand, clip A represents the position of the interviewee, clip C is still true to the position, but clip B gives a wrong impression. Therefore, a side effect of a finer granularity is that clips can be taken out of context and misrepresent what was intended. An automatic generation approach needs to encode context information to present a clip in order to avoid unintentional misunderstandings, as would happen if only clip B was shown, rather than clip A, to represent the interviewee's position.

Furthermore, the audio track must be properly segmented so that the clip does not sound strange to the viewer [7]. This requires starting and ending the clip at appropriate points of the interviewee's answer, respecting word boundaries as well as the intended meaning of the sentences. For example, in Fig. 1, neither Clip B nor Clip C contains the "but" in between the two sentences, since a clip starting or ending with "but" would give the viewer the impression that a part of the answer was left out by mistake. Nevertheless, the semantics associated with "but" cannot be lost: although they are contained in Clip A, they also need to be encoded in the context information associated with Clip B and Clip C. Context information can be provided by determining the role each of the interviewee's statements plays in building the argument.

4.1.4. *The Toulmin model*

Analyzing the example in Fig. 1, it is clear that not all sentences an interviewee says have the same weight when expressing her position. How important a part of an argument is can be determined using an argument model, such as the Toulmin model [24]. This argumentation model is commonly used in argumentation studies to diagram the domain independent way an argumentation works. This model is not concerned with the

soundness of the argumentation but describes the general structure of rational argumentation, by identifying the different discourse parts used to make a claim and their role. According to Toulmin, an argument can be broken down into the following functional components:

– a claim is a statement being argued for, the conclusion of the argument, concerning a potentially controversial issue, for example, "*war is the right solution*";
– the data [8] are facts or observations about the situation under discussion, and are the basis for making the claim, for example, "*we have been attacked*";
– a warrant is the chain of reasoning that connects the data to the claim, usually based more on common sense than on strict logic, for example, "*if you are attacked, then you must react with violence*";
– a backing is the theoretical or experimental foundation that justifies the warrant, for example, "*waging a war prevents future attacks, since a war damages the opponent and causes a reduced capability to attack again*";
– a qualifier expresses the degree of certainty for the claim, for example "*always*" or "*sometimes*";
– rebuttals are possible exceptions to the claim, which can be used as counterarguments to it. They can be:
  · concessions that contradict but are less strong than the claim, for example, "*even though war kills innocent people*"
  · conditions that, if true, could invalidate the claim, for example, "*as long as no innocent people are killed*".

In our model, we do not use the qualifier, since its function is equivalent to the statement's modifier (see section 4.1.1), and it would duplicate functionality. This modification is also used by other approaches [2,18]. The adapted Toulmin model is represented in Fig. 2. Using the Toulmin model to annotate the example in Fig. 1, we see that the statement expressed in clip B is a concession, which only expresses a concern of the speaker, while clip C is the claim, the point the speaker wants to make. Therefore, Toulmin can explain why presenting Clip B (i.e. the concession in the interviewee's argument) is misrepresenting the real position of the interviewee. Context information is therefore provided by encoding Toulmin's role together with the statement expressed in a clip.

---

[7] video must also be properly segmented, but in interviews the video track is sufficiently static with respect to the audio track, and a proper audio segmentation usually yields also a proper video segmentation.

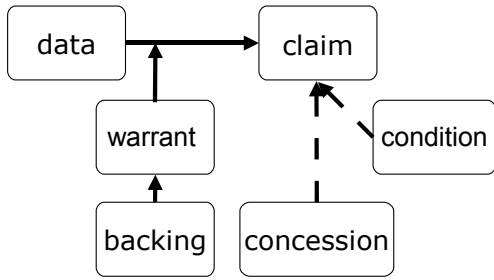[8] In literature the data are also called the grounds.

Figure 2. The adapted Toulmin model (without the modifier). Dashed lines indicate rebuttals.

4.2. *Generating a documentary*

As we saw in section 3.3, a *Semantic Graph* relating the concepts and the media items contained in the repository can support the process of presentation generation, i.e. it can provide the story space within which particular presentation (or documentaries) can be created. In this section we explain how the generation process first creates such graph using our annotation schema (statements plus thesauri) and then uses it to generate documentaries. A *Semantic Graph* is in general a graph whose nodes represent concepts and whose edges the relations among them. In our case the nodes are annotations associated with the media items in the repository (i.e. the statements), and the edges are argumentation relations of two possible polarities, either positive (i.e. supports) or negative (i.e. contradicts). The generation process uses these argumentation relations to build arguments according to the rhetorical form described in section 2. The semantic graph needs to be dynamically created each time an annotated media item is added to the repository, by inferring relations among the statements.

The process that creates the semantic graph is composed of two sub-processes, the first generating new statements and the second linking them. The rationale behind this choice is that decomposing the process into two steps allows to better pinpoint how statements and terms in the thesauri contribute to the graph creation. This information can be used to improve the quality of the generated semantic graph ([5]). We describe these two sub-processes in detail in the following two sections. During the discussion we use as examples terms and relations from Tables 1, 2 and 3.

4.2.1. *Generating new statements*
The aim of the first sub-process is to generate, for each existing statement s_n that annotates a video clip in the repository, the set of all possible related statements

Statements			Transformations	Link
subject	modifier	predicate		
war	not	effective	*Generalization* s	SUPPORTS
economic aid	not	effective	*Opposite* s	CONTRADICTS
bombing	not	waste	*Opposite* p	CONTRADICTS
war	no mod	effective	*Generalization* s	CONTRADICTS
			- *Opposite* m	
peace	not	effective	*Generalization* s	CONTRADICTS
			- *Opposite* s	
bombing	no mod	waste	*Opposite* p	SUPPORTS
			- *Opposite* m	
war	not	waste	*Opposite* p	CONTRADICTS
			- *Generalization* s	
military actions	no mod	effective	*Generalization* s	CONTRADICTS
			- *Opposite* m	
			- *Similar* s	
economic aid	no mod	effective	*Generalization* s	SUPPORTS
			- *Opposite* s	
			- *Opposite* m	
bombing	no mod	useless	*Opposite* p	SUPPORTS
			- *Opposite* m	
			- *Similar* p	
war	not	useless	*Opposite* p	CONTRADICTS
			- *Generalization* s	
			- *Similar* p	

Table 4

Example of statements generated from s:bombing m:not m:effective using transformations on subject, modifier and predicate with terms and relations from Tables 1, 2 and 3 (*Similar* s means apply relation *Similar* to the subject, and so on). The last column reports the type of link to the original statement in terms of the argumentation relations SUPPORTS and CONTRADICTS.

R^{s_n}, regardless of whether these generated statements correspond to a video clip in the repository [9]. The end result is a semantic graph containing all potential relations (edges) among statements, i.e. all potential arguments that can be generated given the annotations. Some of these statements do not correspond to any media item in the repository. The second sub-process then selects only the statements that correspond to video clips in the repository. This process checks, for each generated

---

[9] The set of all possible related statements grows linearly with the number of statements, which makes our approach scalable to larger repositories (see [4], p. 103-104).

statement $s_g \in R^{s_n}$, whether s_g is equal to an existing statement s_m, i.e. that the subject, modifier and predicate of the two statements are the same. If this is the case, s_m is related to the initial statement s_n, and the second sub-process links s_n and s_m together, otherwise s_g is discarded.

The input to the first sub-process is the set of statements contained in the repository (existing statements). New statements are generated by replacing the terms in the existing statements with related terms contained in the corresponding thesaurus. The rationale for this is that the relation between two terms in the thesaurus can be used to infer the relation between two statements that contain these terms, as we show in section 4.2.2.

We describe now how to generate new statements from an existing one. For each existing statement, the first sub-process retrieves the subject, modifier and predicate. Each new statement is generated by replacing either the subject, the modifier or the predicate of the original statement with a related term. The thesaurus defines whether two terms are related, and with which relation: either *Similar*, *Opposite*, *Generalization* or *Specialization* (section 4.1.2). At this stage, each new statement is equal to the original one with the exception of one term, i.e. either the subject, the modifier or the predicate. For example, let us assume that the original statement is s:bombing m:not p:effective. The term *bombing* is *Opposite* to the term *economic-aid* and *Generalization* to the term *war* in the thesaurus (Table 4). The process thus is able to generate the following two new statements: s:war m:not p:effective and s:economic-aid m:not p:effective. Replacing one term constitutes one round of transformations. The same process is applied again to the generated statements. At each transformation round, the difference from the original statement increases: at the n-th round the new statements have been obtained by replacing n times terms from the original statement. Each term used as the subject, the modifier or the predicate in a generated statement is related through one or more thesaurus relations to the term in the corresponding part of the original statement. We limit heuristically the number of transformations to three, as discussed in the following section.

Table 4 shows examples of new statements that can be generated from the statement s:bombing m:not m:effective, using the thesauri shown in Tables 1, 2 and 3, with up to three rounds of transformations. Transformations are represented in the second column with two terms (e.g., *Opposite s*), the first being the name of relation in the thesaurus, which relates the replaced term to its replacement (in the example *Opposite*), and the second which statement part has been replaced (in the example the

subject). The statement in the fifth row, for example, s:peace m:not m:effective, has been generated from the statement s:bombing m:not m:effective using two rounds of transformations. First, by replacing the subject *bombing* with *war* (since *bombing Generalization war* in Table 1), giving s:war m:not m:effective (first row in Table 4). Then, replacing again the subject *war* with *peace* (since *war Opposite peace* in Table 1), giving s:peace m:not m:effective. In this particular example, the subject of the statement has been replaced twice, but at each round, any of the statement parts can be replaced.

Since generated statements are composed of terms that are related to the terms of the original statement, the statements are also related. The generated statements can be considered as being in the "semantic neighborhood" of the original statement. Generated statements represent the semantic "mutations" of the original statements based on the relations provided in the thesaurus. Not all the "mutations" correspond to video clips in the repository.

4.2.2. *Linking statements*

The goal of the second sub-process is to establish which statements should be linked together and how. To verify whether a generated statement is equal to an existing statement, this sub-process searches for it among the annotations in the repository. A generated statement that is found in the repository generates a hit. A generated statement can generate no hits if there are no video clips annotated with that statement, or one or more hits if one or more video clips are annotated with the same statement.

Once it has been established that two existing statements should be linked, the link type must be determined. The link must be either SUPPORTS or CONTRADICTS. We assign the link type based on the transformations used by the first sub-process to get from the original statement to the generated one. To map from transformations applied to either SUPPORTS or CONTRADICTS links, we use the following criterion: if the statement is derived using no or an even number of *Opposite* relations, we assume that the link is SUPPORTS, otherwise the link is CONTRADICTS. For example, the statement s:economic-aid p:effective (9th row in Table 4) has been derived from s:bombing m:not m:effective with three transformations: *Generalization s*, *Opposite s*, *Opposite m*. We therefore conclude that statement s:bombing m:not m:effective SUPPORTS s:economic-aid p:effective.

This inferencing method is therefore based on a simple logic, which defines the relations *Similar*,*Generalization* and *Specialization* as always yield-

ing a SUPPORTS link, while the *Opposite* one as always yielding a CONTRADICTS link, with the assumption that if statement$_A$ CONTRADICTS statement$_B$ and statement$_B$ CONTRADICTS statement$_C$, then statement$_A$ SUPPORTS statement$_C$. This logic cannot guarantee to produce always meaningful links. On the other hand, even a more complex logic would yield some unreliability, since we operate under an open-world assumption (as Nack [17] concludes). Intuitively, the more transformations used to derive a statement, the less we can rely on the conclusion. On the other hand, the more transformation rounds are used, the more new statements are generated, the more the probability to find a link increases. In [4], pp. 103-105 we show that the number of links between statements generated at each transformation round forms a Gaussian curve, which peaks at 2 and then slowly decreases. After a certain point it seems that the generated statements become semantically too far removed from the original content of the repository, as if too many manipulations have led to statements that make no sense or make no sense in the domain of the repository. Limiting the transformations to 3 represents a good trade-off between results obtained, time required and link validity.

The end result of this phase is a semantic graph where the nodes are the statements and the edges are either SUPPORTS or CONTRADICTS links. Since each statement is associated to a media item, the corresponding media items are also linked by either SUPPORTS or CONTRADICTS links. The generation process can thus use this data structure to compose arguments.

4.2.3. *Composing arguments*

The story space we built in the previous section can be used to compose a single argument. An argument is based on a single interview segment, complemented by statements contained in other interview segments. The composition is based on the relations between the interview's statements and other statements contained in the semantic graph.

In order to compose statements into an argument, there must be a relation between them that motivates the composition. If they are related, two statements either support each other or they contradict each other. The generated semantic graph can be used to counterargue or support an argument to be presented. To counterargue an interview we state a conclusion opposite to the given one or, using the Toulmin model, a statement that contradicts the claim. Since in an argument each part is encoded as a statement, rebuttals for a particular argument are all the statements that have a CONTRADICTS

link in the semantic graph to the statement representing the argument's claim. Analogously, to support an interview we select all the statements that have a SUPPORTS link to the statement representing the argument's claim. Moreover, in the Toulmin model, the data, warrant and the backing support the claim, while the concessions and the conditions counterargue it. Counterarguing an argument can therefore be done in two more ways: either contradicting a part that supports the claim, or supporting a part that counterargues the claim. The first case corresponds to selecting all the statements that have a CONTRADICTS link to any of the statements supporting the claim, while the second corresponds to selecting all the statements that have a SUPPORTS link to the statements counterarguing it. Analogously, the statements supporting an argument are all the statements that have a SUPPORTS link to any of the statements supporting the claim plus all the statements that have a CONTRADICTS link to the statements counterarguing it.
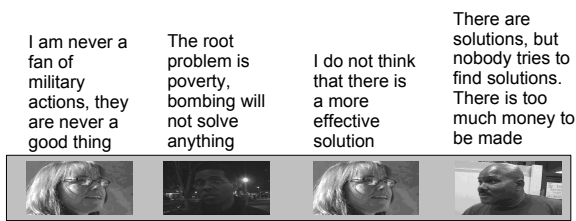
Once the video clips forming an argument have been selected, they need to be edited into a sequence to be presented to the viewer. In the editing phase, clips are ordered in a linear sequence and joined together using either cuts or transitions. Ordering requires that the initial structure, consisting of the interview segment (modeled with the structure of Toulmin) and the corresponding supporting and/or counterarguing clips, is transformed into a linear sequence. We show some examples of editing while discussing an implementation of the model in the next section.

4.3. *Vox Populi*

Vox Populi [10] is an implementation of the video generation model that we made to demonstrate the model's functionality. Due to space limitations, some of the features contained in Vox Populi have not been described in section 4, and can be found in [4], such as editing continuity rules and feedback indexes for the annotator on the quality of the annotations. We test our approach on material from **Interview With America** (IWA) [11], which is an online documentary shot by a group of independent amateur documentarists on the events happening after the terrorist attack on September 11th 2001. The annotations cover one hour of video footage, containing 15 interviews, 60 interview segments and 120 statements, composed with 155 terms from the thesaurus.
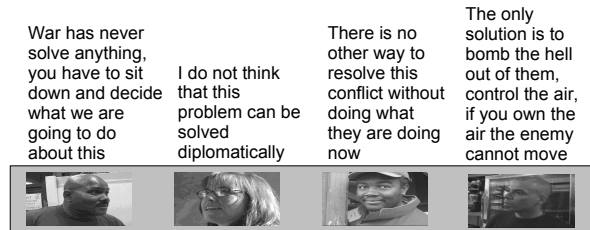
---

[10] http://www.cwi.nl/~media/demo/VoxPopuli/
[11] http://www.interviewwithamerica.com/
documentary.html

| I am never a fan of military actions, they are never a good thing | The root problem is poverty, bombing will not solve anything | I do not think that there is a more effective solution | There are solutions, but nobody tries to find solutions. There is too much money to be made |

Position: war in Afghanistan – For
Point of View: Propagandist – Create Clash

Figure 3. —Generated documentary with Interviewee "Lawyer in Harvard"and clashing interviewees of "Race" = "Black".

| War has never solve anything, you have to sit down and decide what we are going to do about this | I do not think that this problem can be solved diplomatically | There is no other way to resolve this conflict without doing what they are doing now | The only solution is to bomb the hell out of them, control the air, if you own the air the enemy cannot move |

Position: war in Afghanistan – Against
Point of View: Propagandist – Create Clash

Figure 4. —Generated documentary with Interviewee "Black shop owner Stanford"and no limitations on clashing interviewees.

| There is no other way to resolve this conflict without doing what they are doing now | I am never a fan of military actions, they are never a good thing | We have to fight, we have to fight | If you have a good point, you should be able to tell people about it |

Position: war in Afghanistan – Against
Point of View: Propagandist – Create Clash

Figure 5. —Generated documentary with Interviewee "Cameroun Parking Guard at Stanford"and clashing interviewees of "Gender" = "Female".

Using Vox Populi, the viewer can request documentaries, by specifying first the subject and then the point of view. To choose the subject, the viewer can select one or more **questions**. All those clips are retrieved where an interviewee replies to this/these question(s). Alternatively, she can select one or more **positions**. We model positions using two values, the subject, which is a controversial issue such as "*war in Afghanistan*", and the interviewee's attitude with respect to the subject, which can be "*for*", "*against*" and "*neutral*". All the clips where this/these position(s) is/are expressed are retrieved. She can also select one or more **interviewees**. All the clips where this/these interviewee(s) is/are shown are retrieved. If the viewer selects more options, only clips corresponding to all the options are selected.
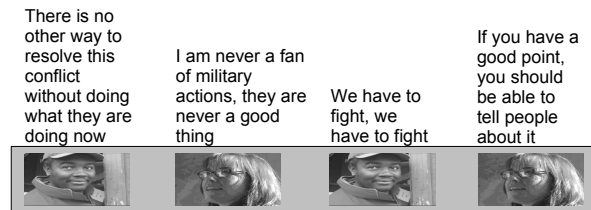
After the content for the documentary has been specified, the viewer can select the **Point of View** she wants the documentary to have. A rhetorical documentary can have three different points of view: **Propagandist - Create clash**, **Propagandist - Create support** and **Binary Communicator**. A propagandist presents only one position or makes one position look stronger than the other, while the binary communicator strives to present contrasting positions with equal strength. To implement a propagandist point of view, Vox Populi selects only clips counterarguing (in the *Propagandist - Create clash* case) or supporting (in the *Propagandist - Create support* case) the chosen interview, as explained in section 4.2.3. In the *Binary Communicator* case, Vox Populi selects both counterarguing and supporting clips. Furthermore, the viewer can select the social categories of the interviewees taking part to this rhetorical debate, i.e. *age*, *education*, *employment*, *race*, *religion*, and *gender*. In this way we can have, for example, a documentary where an initial interview is counterargued by highly educated white people and supported by non educated afro-american people.

In Fig. 3 we show a documentary generated specify-ing for the **Position** "war in Afghanistan - For" and as **Interviewee** "Lawyer in Harvard" (the woman shown on the left). The **Point of View** is "Propagandist - Create Clash" and the counterarguing group is selected to have **Race** "Black". In Fig. 4 the Black shop owner in Stanford expresses a position against the war in Afghanistan and he is counterargued by three people. In Fig. 5 the Cameroun Parking Guard at Stanford expresses a position for the war in Afghanistan, and he is counterargued by the Lawyer (who shares actually his position) using two fragments taken out of context.

As well as for IWA, Vox Populi has been used in two other projects, namely VJ Cultuur [12], which aims at describing the work of VJs with respect to other art disciplines and the reciprocal influences between existing visual arts and VJ, and Passepartout [13], which investigates the technical challenges and new forms of human-computer interactions in broadband home environments. Although the goal and the domain of both projects are different from IWA's, we have used this experience to extend and improve Vox Populi's implementation.

5. Conclusions and future directions

In the paper we present an alternative way of authoring documentaries, which a documentarist might decide to use instead of the traditional documentary making process. The use of our approach is not necessarily an alternative to traditional documentary making: our system could also be used to suggest interesting editing possibilities, based on different points of view. The documentarist might adopt, or expand on, generated sequences to create a final static version, or she could use the system as a way of browsing the content of the footage, instead of using transcripts and logs.

In our approach the content of the repository cannot be foreseen beforehand, or, in other words, the model needs to operate under an open-world assumption rather than a closed-world one. Since our annotations do not capture all the information contained in video clips, some generated combinations of clips may make no sense or be either unqualified or offensive, which clearly produces the opposite effect to the intended one. On the contrary, in systems operating under a closed-world assumption, such as Terminal Time [14], the material is created specifically for the automatic generation task. This allows complete control over the content and the possible sequences that can be generated. Analogously to what Nack [17] concludes, some unreliability is the price one has to pay for operating in an open-world setting.

An obstacle to the adoption of our approach is the effort needed to annotate the material. The annotations we require are considerably more complex than keywords and have to be done manually since they cannot (yet easily) be obtained automatically from video and audio processing methods. Obtaining annotations automatically would greatly facilitate the adoption of approaches such ours. Further research is needed in this field, and we are looking at speech recognition techniques, as well as at video segmentation techniques.

The role of **non-verbal visuals** in communicating a message or strengthening a verbal message needs to be further researched, and more theories relating video to viewer reactions need to be developed. Documentarists stress the importance of providing visual evidence and a visual story which complement what is said in the documentary. Further research is needed to establish how and when to use visuals, and also to understand the effects of combining non-verbal visuals with auditory verbal messages, e.g. using counterpoint editing. As a side effect of this, we will gain a deeper understanding of interpreting media, and the ways media (and viewers) can be manipulated, of which the work presented in the paper is an example.

6. Acknowledgements

References

[1] H. Beck, H. S. Pinto, Overview of Approach, Methodologies, Standards, and Tools for Ontologies, The Agricultural Ontology Service (UN FAO).
URL http://www.fao.org/agris/aos/Documents/BackgroundPaper.pdf

[2] T. J. Bench-Capon, Specification and Implementation of Toulmin Dialogue Game, in: Proceedings of JURIX 98, 1998.
URL http://www.jurix.nl/pdf/j98-01.pdf

[3] S. Bocconi, VOX POPULI: Automatic Generation of Biased Video Sequences, in: First ACM Workshop on Story Representation, Mechanism and Context, 2004.
URL http://www.cwi.nl/~media/publications/srmc04-workshop.pdf

[4] S. Bocconi, Vox Populi: generating video documentaries from semantically annotated media repositories, Ph.D. thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, also available on http://www.cwi.nl/~media/group-www/theses/StefanoBocconi06.pdf (November 30, 2006).
URL http://www.cwi.nl/~media/group-www/theses/StefanoBocconi06.pdf

[5] S. Bocconi, F. Nack, L. Hardman, Supporting the Generation of Argument Structure within Video Sequences, in: Proceedings of the sixteenth ACM Conference on Hypertext and Hypermedia 2005, 2005.
URL http://www.cwi.nl/~media/publications/ht2005.pdf

[6] D. Bordwell, K. Thompson, Film Art: An Introduction, 7th ed., McGraw-Hill, 2003.

[7] C. Cleary, R. Bareiss, Practical methods for automatically generating typed links, in: Proceedings of the the seventh ACM conference on Hypertext, Bethesda, Maryland, United States, 1996.
URL http://portal.acm.org/citation.cfm?id=234832&coll=Portal&dl=ACM&CFID=40687806&CFTOKEN=41442351

[8] M. Crampes, Auto-Adaptive illustration through conceptual evocation, in: Proceedings of the 2st ACM international conference on Digital libraries, ACM, Philadelphia, Pennsylvania, United States, 1997.
URL http://doi.acm.org/10.1145/263690.263829

[9] Dublin Core Community, Dublin Core Element Set, Version 1.1 (2003).
URL http://www.dublincore.org/documents/dces/

[10] J. Geurts, S. Bocconi, J. van Ossenbruggen, L. Hardman, Towards Ontology-driven Discourse: From Semantic Graphs to Multimedia Presentations, in: D. Fensel, K. Sycara, J. Mylopoulos (eds.), Second International Semantic Web Conference (ISWC2003), Springer-Verlag, Sanibel Island, Florida, USA, 2003.
URL `http://www.springerlink.com/openurl.asp?genre=article&issn=0302-9743&volume=2870&spage=597`

[11] B. Hampe, Making documentary films and reality videos: a practical guide to planning, filming, and editing documentaries of real events, Henry Holt and Company, 1997.

[12] C. Lagoze, H. V. de Sompel, The Open Archives Initiative: Building a low-barrier interoperability framework, JCDL2001.
URL `http://www.openarchives.org/documents/oai.pdf`

[13] S. Little, J. Geurts, J. Hunter, Dynamic Generation of Intelligent Multimedia Presentations through Semantic Inferencing, in: 6th European Conference on Research and Advanced Technology for Digital Libraries, Springer, Pontifical Gregorian University, Rome, Italy, 2002.
URL `http://www.cwi.nl/~media/publications/ecdl2002.pdf`

[14] M. Mateas, Generation of Ideologically-Biased Historical Documentaries, in: Proceedings of AAAI 2000, 2000.
URL `http://www-2.cs.cmu.edu/~michaelm/publications/AAAI2000.pdf`

[15] S. Melnik, S. Decker, Wordnet RDF Representation, http://www.semanticweb.org/library/ (2001).
URL `http://www.semanticweb.org/library/`

[16] M. Murtaugh, The Automatist Storytelling System, Ph.D. thesis, Massachusetts Institute of Technology (1996).
URL `http://alumni.media.mit.edu/~murtaugh/thesis/`

[17] F. Nack, AUTEUR: The Application of Video Semantics and Theme Representation in Automated Video Editing, Ph.D. thesis, Lancaster University (1996).
URL
`http://homepages.cwi.nl/~nack/thesis.pdf`

[18] S. Newman, C. Marshall, Pushing Toulmin too far: learning from an argument representation scheme, Xerox PARC Technical Report SSL-92-45, 1992. (1992).

[19] M. Rabiger, Directing the Documentary, Focal Press, 1998.

[20] W. Sack, Coding News And Popular Culture, in: The International Joint Conference on Artificial Intelligence (IJCA'93). Workshop on Models of Teaching and Models of Learning, Chambery, Savoie, France, 1993.

[21] K. Schwarz, T. Kouwenhoven, V. Dignum, J. van Ossenbruggen, Supporting the decision process for the choice of a domain modeling scheme, in: Formal Ontologies Meet Industry, Verona, Italy, 2005.
URL `http://www.cwi.nl/~media/publications/fomi_position_paper_june2005.pdf`

[22] S. B. Shum, E. Motta, J. Domingue, SchollOnto: an Ontology-Based Digital Library Server for Research Documents and Discourse, International Journal on Digital Libraries 3 (3).
URL `http://kmi.open.ac.uk/projects/scholonto/docs/ScholOnto-IJoDL-2000.pdf`

[23] R. Stam, R. Burgoyne, S. Flitterman-Lewis, New Vocabularies in Film Semiotics: Structuralism, Post-Structuralism, and Beyond., Routledge, 1992.

[24] S. Toulmin, R. Rieke, A. Janik, Introduction to Reasoning, 2nd ed., MacMillan Publishing Company, 1984.

[25] M. van Assem, M. R. Menken, G. Schreiber, J. Wielemaker, B. Wielinga, A Method for Converting Thesauri to RDF/OWL, in: Proceedings of the Third International Semantic Web Conference (ISWC'04), No. 3298 in Lecture Notes in Computer Science, Springer, Hiroshima, Japan, 2004.
URL `http://www.cs.vu.nl/~mark/papers/Assem04a.pdf`