Invited Paper

Explain This to Me!

A Study on Automatic Recompile of Broadcast News Video

Ichiro Ide (member)^{†,††}, Frank Nack^{†††}

Abstract This paper addresses a framework that facilitates the semi-automated authoring of already edited new stories available in the repository of a news corporation or a public broadcast video archive. The newly generated video explains the chronological development of a current event, such as the resignation of a Prime Minister. The aim is to facilitate a journalist with an audio-visual body based on which he/she can finalize the explanatory piece. The framework introduces techniques that exploit demoscopic data in form of polls for the development of the general story outline; the automatic retrieval of relevant material by using a combination of event templates and automatic news summarization over topic threads; and the generation of the final video by applying a set of trimming rules. Example generations are presented and discussed, and an outline of future work is presented.

Key words: Multimedia authoring, topic, event, automated video editing, story generation

1. Introduction

A news program on TV is regularly scheduled and reports current events. They typically follow a particular format, i.e. they start with the head story, followed by international news, national news, sports, and the weather, where each part forms a series of individual stories that are presented by an anchor person. There are in general, two news formats, the central news show of around 15 minutes and a longer version of around 30 minutes, which provides more in-depth information. In particular, the longer format includes live material, e.g. recorded interviews by field reporters, expert opinions, and demoscopic polls, or editorial content, i.e. short background information that explains a current event, for example about the resignation of a national leader.

In particular, the editorial content for short background information to a headline of the program requires a complex and resource demanding production process, structured for very rapid assembly of material from very diverse sources¹⁶⁾. This type of explanatory story needs to explain within a given time constraint

that fits the overall news structure, what the government has achieved or not, and what was the final reason that lead to its fall. The making of this piece requires that the most significant events during the government are selected, where two aspects are important. First, the events need to speak to the society's memory so that the audience can reflect about the presentation. Second, the stories represent the information bias of the news corporation as any news is presented with some bias. From the point of view of a responsible editorial journalist, the most complex aspects of this authoring process is to identify the significant events for a given topic and present it according to the broadcaster's ideological presentation style. Usually the journalist will make use of his/her own memory shaped by his/her work focus, to establish the story-line and develop the explanatory voice-over. During the process, he/she will frequently access the news corporation's repository, to ensure that the relevant content is available in the appropriate format. The whole process needs to be performed on a very short notice.

In this paper, we present work that is directed towards the application of IT support for the outlined production process, including the automation of the message construction, query, asset creation, and packaging processes¹⁰⁾. The aim is to facilitate the journalist with an audio-visual start-up video that chronologically presents crucial events where the order explains the topic under investigation. The goal of the research is to establish a framework of techniques that facilitates semi-automated authoring of explanatory audio-visual

This paper is an extended version of the paper presented at the 19th ACM International Multimedia Conference on November 30, 2011.¹⁷⁾

Received December 1, 2012; Revised February 12, 2013; Accepted March 4, 2013

[†] Graduate School of Information Science, Nagoya University (Nagoya, Japan)

^{††} National Institute of Informatics, Research Organization of Information and Systems (Tokyo, Japan)

^{†††} ISLA, Informatics Institute, University of Amsterdam (Amsterdam, The Netherlands)

news topics in a retrospective style for the domain of politics based on already edited news stories available in the repository of the news corporation. The relevance of the events chosen to explain the causal chain results from a connection of the story events to positive or negative developments of tracking polls that cover the period in which the topic under consideration occurred. In that way, the proposed authoring framework makes use of the collective memory of the viewers, as the used clips function as “triggers” to well-known events that have been popularized already on TV. The choice and order of events represent the general practice of the broadcasting corporation to take advantage of a topic by selecting some story events and sacrificing others to better present the particular point of view of the news organization. Thus, the presented framework includes techniques that establish the often expected aim for objectivity but also facilitates a biased analysis of the topic under investigation. On both aspects, namely, content selection and its final presentation, we aim to demonstrate and assess the applicability and acceptability of this framework in a news production environment.

In the following, we first define a number of key attributes of our framework, and then describe its basic elements. These are the exploitation of poll data for the development of the general story outline; the automatic retrieval of related material by using a combination of event templates and automatic news summarization over topic threads; and the final generation of the media material through automated editing. Finally, we assess our achievements thus far by providing and discussing three generation examples, and provide an overview of further work.

2. Related work

There are in particular, three fields in multimedia research that are relevant for the approach described in this paper; news video semantics, semi-automated multimedia authoring, and event representation.

There is a large body of work within multimedia related to news videos. The predominant amount of work focuses on helping with semantic video retrieval¹²⁾²⁵⁾, which is relevant for the detection of topics in non-annotated material. Other works cover the aspect of video summarization⁵⁾⁷⁾²⁴⁾, which address the problem of distilling appropriate videos from large corpora. Work on topic clustering points into the same direction¹⁸⁾, only that here the interest is not so much to

support the retrieval of video but rather to extract the connection between news stories over time. The problem with these approaches is that they mainly focus either on the structural extraction of a single news show or try to concatenate the different threads of one news story over time. The combination of different story threads is so far not addressed, though the work of Ide et al. on “mediaWalker”¹³⁾ is the closest reaching into this direction. Its aim is to establish and exploit the chronological semantic structure in a large-scale broadcast news video archive. The results of the mentioned work, especially work on the message construction and asset querying processes, provide us with the basis for the part in our framework that aims at detecting explanatory events of poll behaviors.

Multimedia authoring, seen as the idea of allowing an endless montage of audio-visual assets into new items, inspired a great deal of research in computer environments that embody mechanisms to interpret, manipulate or generate audio-visual media, summarized by Bulterman and Hardman⁴⁾.

Davenport and Murtaugh⁶⁾ proposed a model for the “evolving documentary” to be applied on a large audio-visual hypermedia system that covers the public works on rebuilding the Central Artery and the project’s impact on surrounding neighborhoods. The idea behind the “evolving documentary” is that a story rarely includes all relevant material but that only material that best fits the viewer’s information interest is selected and sequenced. As argumentative continuity plays a key role in a documentary, Davenport and Murtaugh introduced description feedback, which establishes a numerical relation between the viewer’s experience model and the material’s context model created by the author of the hypermedia space.

Nack¹⁵⁾ presented “AUTEUR”, a system that fully automatically generated humorous, non-verbal video sequences. The aim of the research was to define techniques to automatically assemble video sequences that realize a given theme, ultimately to be used to assist in the presentation or interpretation of video. The work represents a planner based approach to the application of video semantics and theme representation to the automated editing of visual stories at the level of events. Nack introduced representational structures for the semantic, temporal and relational features of video; representations and strategies to support automated editing, in combination with a simplified model of the film editing process; and representations for narrative structures

such as actions, events, and emotional codes.

Systems as the virtual storyteller system²⁷⁾ and “AConf”¹⁹⁾ are examples for approaches that seek to generate full narratives for entertainment. Systems like “Topia”²⁰⁾ make use of narrative devices, such as sequencing, emphasis and omission, to incorporate discourse into Web technology and the Web experience.

Nack and Putz¹⁶⁾ considered the automated and semi-automated annotation of audiovisual media in a new type of production framework, “A4SM” (Authoring System for Syntactic, Semantic and Semiotic Modelling). They presented the architecture of the framework and outline the underlying XML-Schema based content description structures. In addition, they presented a metadata generating camera, a commenting tool for journalists to be used during news production, and an editing suite to be used on location and with those demonstrated how video material can be annotated in real time and how this information can not only be used for retrieval but also during the different phases of the production process itself.

Bocconi and Nack³⁾ presented an experimental rhetoric engine “Vox Populi” that utilizes the IWA (Interview with America) repository of video interviews with U.S. residents on the events happening after 9/11 to generate biased video documentaries. They defined a simple rhetoric annotation system, based on the Toulmin model²⁸⁾, which facilitates the identification of statements and the analysis of the rhetoric structure in which the statements are placed, to encode the verbal information contained in the audio channel. The annotations formed the basis for the generation engine.

Scherp and Boll²¹⁾ developed a context-aware smart multimedia authoring tool “xSMART”. xSMART is a semi-automatic authoring tool that integrates the targeted user context into the different authoring steps and exploits this context to guide the author through the content authoring process. The design of xSMART facilitates its customization to the requirements of a specific domain by domain-specific wizards.

Hargood et al.¹¹⁾ investigated the use of theme for the generation of photo narratives and found that a thematic model is capable of successfully connoting themes within these narratives.

All these works address the various steps within multimedia production¹⁰⁾, except Nack¹⁵⁾ always assuming that there is an author involved in the process. In our framework, we also assume an author in the production loop, in particular for the making of the voice-over, but

Fig. 1 Example of a topic thread structure originating from a story under consideration.

we aim for an automated production process that addresses the parts, message construction (establish the story-line) and packaging (trim the retrieved material for the final output)¹⁰⁾.

In recent years, substantial work has been provided for the modeling of multimedia events¹⁾²³⁾²⁹⁾. The proposed models are relevant as they model time and space, objects and persons, as well as causal and correlative relationships between events for particular domains. In that way, they represent the basic questions about any particular event from the point of view of news, namely—who, what, when, where, why, and how (5W1H, see in particular the discussion by Scherp et al.²²⁾). We follow a similar path but represent contextualized political events, in form of collections of particular useful document sources and the particular way how they should be interpreted so that the system can establish the potential main events. We also make use of a combination of topic-based and template-based content identification techniques so that the material for the established events can be retrieved.

3. Proposed framework

In this section, we explain the key steps of the framework we propose for the semi-automatic generation of videos that explain a current news event in a retrospective view. As we use particular terms throughout the paper, we first define those. We base the definition of these terms on two sources, namely the Topic Detection and Tracking (TDT) workshop series organized by NIST⁸⁾ and the work by Ide et al.¹³⁾. The definitions of the first three adopt TDT’s view, while the latter two follow those introduced by Ide et al.

- Event: Some incident that occurred at some specific time and place along with all necessary preconditions and unavoidable consequences.
- Story: A topically cohesive segment of news that includes two or more declarative independent clauses about a single event.
- Topic: A seminal event or activity, along with all

Fig. 2 Outline of the generation process.

directly related events and activities.

- Topic thread: A sequence of related stories chained chronologically. It may contain several topics.
- Topic thread structure: A directed graph composed of topic threads terminating/originating from a specified story.

As shown in Fig. 1, a topic thread structure represents the development of news topics along the time-line concerning a story under consideration.

3.1 Outline

The overall structure of our framework covers three essential phases: First, it needs to determine the topics that the retrospective video should cover. Next, it identifies the essential events that need to be covered. Finally, it establishes the final presentational form of the video. Fig. 2 shows the flow of the process.

Note here that our framework mainly covers the generation of the video itself. Amendments certainly need to be performed once the responsible journalist has finalized the voice-over based on the event order established in the video.

For the ease of readability, we will use the resignation of one Prime Minister, namely Shinzo Abe as later described in Fig. 6, as the example of a current news event to be retrospectively analyzed. The assumed bias for the explanations and the example generations (see section 4) is negative. However, a positive* video could be generated by reverting the bias to positive and then

apply the related structures. Detailed explanations on each phase follows.

3.2 Dataset

The video data used in our work comes from the NII-TVRECS archive¹⁴⁾, which consists of a Japanese news show recorded daily for more than eleven years (2,000–hours). The video data is accompanied with closed-caption, and is segmented into stories by the method proposed by Ide et al.¹³⁾.

From this archive, the proposed framework extracts all stories during the period that contains a Prime Minister's name mentioned as a subject in the closed-caption. For this, morphological analysis** and syntactic analysis*** were applied to all closed-caption texts during the period.

3.3 Input

In order to generate an explanatory video, information is required that provides the essential content and presentation constraints. In our framework, we request the following input parameters for a Prime Minister's resignation:

- Topic: The Prime Minister's name N_{PM}
- Topic duration: His/her period in office (D_{inaug} , D_{resig})
- Bias: Positive / Negative
- Output video length: D_{vid_length}

The generated output is a video that covers essential events, which facilitates a retrospective view on the topic. For our example, this means that a video needs to be generated so that it presents a description of a Prime Minister's key actions in an attempt to explain his/her resignation.

3.4 Story selection

Since our system is designed to be applied in the political domain and should produce an explanatory video to be understood by the audience, we make use of demographic polls as an external reference that should establish the key moments to be considered. Key moments in the relation between politics and polls are those where the government either gains a boost (+10%) or a large loss (−10%) between two poll measurements. Critical moments are also those where the opposition becomes more popular than the government. As the presented opinions relate to real events, we try to identify those

* Or, we may have to say “less negative”, since the data we analyzed showed that most news stories are actually “negative”.

** A Japanese morphological analyzer JUMAN developed at Kyoto University was used. It is available for download from <http://nlp.kuee.kyotou.ac.jp/nl-resource/juman-e.html>.

*** A Japanese syntactic analyzer KNP developed at Kyoto University was used. It is available for download from <http://nlp.kuee.kyoto-u.ac.jp/nl-resource/knp-e.html>.

and in that way we make use of the collective memory of the viewers, as the used clips function as “triggers” to well-know events during the time period of the topic that needs to be explained. The duration of the polls follows the time-frame ^{*4} p_i ($i = 1, 2, \dots, P$) with a period of approximately one month between two consecutive monthly poll measurements while a Prime Minister was in office, where the end of the last day of the poll is considered as the boundary.

Stories that become sources of the final output video are selected by a combination of two strategies; topic-based and template-based. The former is designed to automatically extract related stories that can be linked to the patterns of crucial poll descriptions, while the latter describes prototypical events that establish a plausible introduction to and finish of the output video. Details of each strategy follow.

(1) Topic-based story selection

Significant stories are selected in a bottom-up approach, making use of poll patterns that presumably coincide with major events in political news, and utilizing the features extracted from the topic thread structure concerning each story. The selection is performed in the following steps:

- (1) Time-frames p_i ($i = 1, \dots, P$), where potentially significant events may exist, are determined according to crucial poll patterns.
- (2) The amount of events to be required for the video are calculated.
- (3) Stories that cover the major events within each time-frame p_i are estimated according to the clues obtained from topic thread structures.
- (4) The estimated stories are ranked according to sentiments (negative / positive), in order to be selected as a final story that represents the time period.

Details of the four steps follow.

(a) Step 1: Selecting time-frames according to poll patterns

In order to determine the time-frame where a significant event may have occurred, we refer to polls available from an external source*. The polls consist of ratios on the approval (“Yes”) or the disapproval (“No”) of the

Fig. 3 Example of the transition of polls ($N_{PM} =$ “Shinzo Abe”).

government led by the Prime Minister in focus. **Fig. 3** shows an example of the transition of polls for the government of the Japanese Prime Minister Shinzo Abe. The system determines either positive or negative poll pattern and selects poll positions according to the bias strategy selected by the journalist. Each pattern is defined as follows:

• Positive poll patterns

- The “Yes” rate rises above θ_{vp} .
- The “Yes” rate increases more than θ_{cp} per time-frame.
- The “Yes” rate exceeds the “No” rate.

• Negative poll patterns

- The “Yes” rate drops below θ_{vn} .
- The “Yes” rate decreases more than θ_{cn} per time-frame.
- The “No” rate exceeds the “Yes” rate.

The number of significant poll patterns detected is denoted as N_{poll} . Once this number is calculated, it needs to be determined how many events need to be detected to represent a crucial moment.

(b) Step 2: Determining the overall video structure

This step actually covers two aspects of the video generation, namely content identification as well as structure definition. Both parts cannot be separated from each other. The type of video we consider here follows an already established pattern in broadcasting news, which is as follows: The anchor person introduces the topic explained in the video. The video then starts with some introductory video clips followed by a presentation of the poll graph, on which the crucial data point is highlighted. This is followed by clips of events that explain the data. This alteration between poll graph description and explanatory clips is then followed in alternating fashion until the end, where the video con-

^{*4} Hereafter, [TF] will be used as a unit to count the number of time-frames.

* The “Monthly survey on political attitude” by NHK Broadcasting Culture Research Institute was used as the source. It is updated monthly at http://www.nhk.or.jp/bunken/research/yonon/seijiishiki/list_seijiishiki1.html, and also published in “The NHK Monthly Report on Broadcast Research”.

cludes with the topic that needs to be explained.

Applying this general pattern of retrospective explanation, where the retrospective part is formed through a chronological presentation of events, requires that the pattern can be applied within the given time constraints. For that, a calculation is performed that establishes the number of events that can be presented within each part of the video (Intro, Poll blocks, and Final). This depends on the time period needed to be covered (the larger the number of months between selected polls the more events need to be considered) and the cognitive complexity of the potential clips (e.g. the image of a poll graph can be comprehended within $T_{\text{poll}} = 3$ seconds, where an event clip is usually composed of shots with varying content complexity resulting in a length of somewhere between 30 to 50 seconds⁹⁾).

The calculation in our framework is as such: The start point is the overall time of the final video ($D_{\text{vid_length}}$). From that we subtract $N_{\text{poll}}T_{\text{poll}}$, and the remaining time is divided into equal blocks covering the introduction (Intro), the explanatory events (Poll blocks), and the conclusion (Final), where we follow Gregory's setup of shot ratio comprehension⁹⁾. Depending on the remaining time and the complexity of the middle structure, i.e. the number of critical poll parts needed to be explained, a decision is made on the amount of events to be shown for each explanatory part (see 3.4 (2) for the calculation of the values for the Intro and the Final parts). At the end of this calculation, the basic structure of the video is established. It may look like this: {Intro (2), Poll 1, Expl 1 (2), Poll 2, Expl 2 (2), Poll 3, Expl 3 (2), Poll 4, Final (2)}, where the numbers in brackets represent the amount of stories to be collected for each part. Note that this structure does not provide any editing on the material yet. The final editing of the, by then, established clips will be described in 3.5.

We will now estimate which story may contain an event that may well explain a particular poll pattern.

(c) Step 3: Selecting significant stories during the determined time-frame

First, stories that are assumed to contain significant events are selected according to the "topic thread structure" proposed by Ide et al.¹³⁾. This is done regardless to the established bias. Since a topic thread structure represents the flow of topics preceding and succeeding a story under consideration, we considered that analyzing its structure leads to the evaluation of the significance of the story.

Fig. 4 Terminal / original story in a topic thread structure. Stories concerning general structures such as the one shown in Fig. 1 are neither a terminal nor an original story.

From topic thread structures concerning each story in a time-frame, the story is selected as a candidate of significant stories, if either of the following two criteria is satisfied:

- Original / terminal story: The story in-focus terminates or originates a topic thread structure (Fig. 4).
- Density: More than θ_{da} stories appear along the topic thread structure within θ_{dp} days before / after the story.

The established set of potential stories now needs to be organized so that the amount of events calculated in Step 2 can be extracted.

(d) Step 4: Ranking the selected stories based on sentiment analysis

Finally, sentiment analysis is applied to all the stories selected in the previous step. It is performed in the following manner, which yields a normalized value in the range of $[-1.0, 1.0]$:

- (1) For each term that appears in a story, look-up a sentiment dictionary and assign a p/n value that represents the positiveness / negativeness of the term if it exists.
- (2) Sum-up all the p/n values and normalize it with the number of terms that were assigned a value.

For the sentiment dictionary, Takamura et al.'s dictionary on semantic orientation of Japanese words^{*26)} was used, which is composed of pairs of a word and a value between -1.0 (negative) and 1.0 (positive).

The stories are ranked according to their p/n values, and selected according to the following rules:

- Select the top θ_s stories for a positive poll pattern.
- Select the least θ_s stories for a negative poll pattern.

The number θ_s reflects the value established for the various blocks calculated for the overall video structure. For our example in Step 2, θ_s is 2.

The outcome of the presented four steps is an editing

* Available for download from http://www.lr.pi.titech.ac.jp/~takamura/pndic_en.html.

list that contains the IDs of the stories to be included in the final output video. Note that it might be possible that fewer explanatory events can be established than planned. In such cases, an exception rule will fire that allows a flexible adjustment of the Intro and Final parts with respect to an extension of clips so that the set duration of the video can be kept.

Still missing are the videos for the Intro and the Final parts of the video. Those are established through a template-based story selection process.

(2) Template-based story selection

We make use of templates for the design of the Intro and Final parts of the explanatory video, as they usually portray, according to their rhetoric function, stereotypical events. These events are usually not very much commented on while being shown and only marginally influence the polls and would hence be difficult to detect with the outlined topic-based techniques.

The function of the Intro part is to establish the retrospect setting. Depending on the topic, there are stereotypical events that the potential viewer of the clip will immediately identify. We gather such types of events in the template that is associated with the topic. For the current example, this means that a typical Intro event for a retrospect on a Prime Minister's reign is the "inauguration speech".

The Final part of the video usually provides the resolution to the initial query, meaning that the topic under consideration is actually presented. For our example, this could mean to portray a conflict between a government and opposition in a sequence where the Prime Minister at the end resigns due to the heavy criticism. Our framework reflects on these issues by providing template structures that are related to a topic. We based our design on explanatory videos generated by human editors for the NHK evening news.

Below we describe both the positive and negative events we search for in the presented order. Search here means to look for word sets associated with the event (e.g. the "inauguration speech") to retrieve stories for the Intro and the Final parts. For the Intro part, we make use of the positive part of the template, while for the Final part, we make use of the part that represents best the aimed-for bias. The list of positive events for the topic "resignation" looks like the following (the terms in brackets describe some of the concepts we searched for):

- Inauguration speech (this is always shown) or if not available use the Prime Minister talking in Parlia-

ment [Prime Minister, N_{PM} , inauguration], alternative [Prime Minister, N_{PM} , Parliament]

- If additional material is needed for the Intro part:

- The Prime Minister visits a foreign country (for each nation there might be preferences here for particular countries) [Prime Minister, N_{PM} , China, US, Germany, ...]

- The Prime Minister announces a positive budget [Prime Minister, N_{PM} , budget, positive]

- The Prime Minister presents a positive economical view [Prime Minister, N_{PM} , good economy, improving economy, industry, presentation]

- The Prime Minister wins a prize [Prime Minister, N_{PM} , prize, receive, ceremony]

- Government party wins a by-election [Prime Minister, N_{PM} , election, win]

- Positive statements of the general public

The list of negative events for the topic looks like the following:

- Conflict with the opposition and Prime Minister resigns [Prime Minister, N_{PM} , may resign, opposition party names, leader, party, criticize] and [Prime Minister, N_{PM} , resignation, leaving office]

- If additional material is needed for the Final part (if none of those is already part of the above conflict):

- The Prime Minister has personal or health problems [Prime Minister, N_{PM} , health, hospital, way to hospital]

- Government Party loses a by-election [Prime Minister, N_{PM} , election, loss]

- The Prime Minister or a Minister has juridical problems [Prime Minister, N_{PM} , fraud, party members, ministers]

- The Prime Minister announces a negative budget [Prime Minister, N_{PM} , budget, negative]

- The Prime Minister presents a negative economical view [Prime Minister, N_{PM} , weak economy, high inflation, deflation, industry, presentation]

- Negative statements of the general public

The amount of elements to be presented depends, in accordance to Step 2 in 3.4 (1), on the temporal duration of non-covered poll material. In the proposed framework, stories at the beginning and the end of the entire time-line are selected according to this strategy:

- Inauguration speech

Select from the period $\{p_1, \dots, p_{P_{\text{inaug}}}\}$, the first story that contains the term "inauguration speech" (Number of stories $N_{\text{inaug}} = 1$).

- Opening stories

Depending to the period $\{p_1, \dots, p_{P_{\text{opening}}}\}$ before the occurrence of the first positive / negative poll pattern, N_{opening} stories are selected according to the following criteria:

- $N_{\text{opening}} = 1$ (if $P_{\text{opening}} \leq 2$ [TF])
- $N_{\text{opening}} = 2$ (if 2 [TF] $< P_{\text{opening}} \leq 4$ [TF])
- $N_{\text{opening}} = 3$ (if $P_{\text{opening}} > 4$ [TF])

- Closing stories

Depending on the period $\{p_{P-P_{\text{closing}}+1}, \dots, p_P\}$ after the poll has dropped below θ_{vc} , N_{closing} stories are selected according to the following criteria:

- $N_{\text{closing}} = 1$ (if $P_{\text{closing}} \leq 2$ [TF])
- $N_{\text{closing}} = 2$ (if 2 [TF] $< P_{\text{closing}} \leq 4$ [TF])
- $N_{\text{closing}} = 3$ (if $P_{\text{closing}} > 4$ [TF])

At the end of the complete story selection process, an editing list is established with the selected stories in a chronological order. The identified material still needs to be trimmed so that it fits the overall time constraint set by the journalist.

3.5 Editing for the final output

This part of the framework establishes in a semi-automatic fashion, the final output video. There are basically two steps that need to be performed, namely the generation of the poll inserts and the final trimming of the stories.

(1) Step 1: Generation of the graphical inserts

The graphical inserts serve as the introduction of significant turning points and hence can be understood as punctuation signs that indicate to the viewer that a new context on the way to the resignation is now approaching. Each insert provides a clear identification about the point on the timeline that is now to be covered and why this is important, e.g. the opposition is more popular than the party of the Prime Minister.

Fig. 5 provides an example of such an insert. Generating such inserts automatically can be achieved by using formal descriptions of the various components (e.g. axis, title, highlight circles, etc.) and layout rules for achieving presentational coherence as described by Bateman et al.²⁾. However, they are currently generated manually, though the reimplementation of this component is part of future work.

(2) Step 2: Editing of the story clips and finalization of the video

The final generation step requires that the selected story videos are trimmed so that the given time constraint for the overall length is kept. Based on the allocated material, the length for each clip is calculated as

Fig. 5 Example of a finalized poll insert ($N_{\text{PM}} = \text{"Shinzo Abe"}\text{"}$). The "No" rate exceeds the "Yes" rate in Feb. 2007.

such: When the length of the summarized video is set to T , the length for video clips that should be cropped from each story is decided as

$$T_{\text{clip}} = \frac{T - N_{\text{poll}}T_{\text{poll}}}{\theta_s N_{\text{poll}} + N_{\text{inaug}} + N_{\text{opening}} + N_{\text{closing}}}. \quad (1)$$

The automatic cropping in our framework follows the approach by Nack and Putz¹⁶⁾, which provides a planner for news composition based on 22 rules for the automated clipping of a shot and juxtaposition of shots. In their case, the material to be edited was raw material produced for a particular purpose. Trimming could be achieved in that system by adjusting the video at the beginning or the end. We are, however, confronted with the problem that the selected video stories are already edited and hence the required section is not necessarily found at the beginning or the end of a story clip. We addressed this issue as follows, where the strategies for explanatory parts are slightly different to those used for Intro and Final parts.

(a) Explanatory part

For the explanatory videos, we search in every clip for the topic (here the Prime Minister's name N_{PM} used as a subject) in the closed-caption. The related time-code is taken and then a duration of T_{clip} seconds is clipped from this point into the clip. In case a cut is detected in this part, which means that a new section in the video story has been reached, the sequence is shifted forward. In case the shot only portrays the Prime Minister, i.e. in form of a statement, we make use of the knowledge that our material is already edited, and assume that the real event is covered before. The selected part then starts with the images of the event and resolves with

the statement of the Prime Minister. In cases where the Prime Minister’s name could not be found, we pick the head news, that is the first video clip after the introductory section of the anchorperson and take the first T_{clip} into that part of the story clip. The assumption is that this is the main event to which the poll refers to.

(b) Intro and Final parts

For the clipping of the Intro and the Final parts, we take a slightly different strategy. For the detection of relevant sections of the video story, we make use of the terms associated to the positive or negative events for a topic, as outlined in 3.4 (2). The rule is that for the Intro part, the positive template setting is chosen, while for the Final part, the template setting is chosen according to the aimed-for bias. In both cases, the first event in the template set of events associated with a topic will always be the default.

• Intro part

For the Intro part, the transcript of the selected source video is analyzed for the combination of the terms available in the concept list, for our example “Prime Minister”, N_{PM} , and “inauguration”. The sequence with the highest number of these three terms in the closed-caption is chosen. If it is longer than T_{clip} seconds, it will be clipped from its end on, as it is always easier to cut into an action than terminating on the end of an action. In case the number of events to be covered by the Intro part is larger than one, the list of additional events is sequentially worked through. Identifying suitable points is again based on the distribution of the concepts over the clip where the highest concentration to a particular point in the video determines the entrance point for trimming. The time-code of this point is taken, over which T_{clip} is evenly placed. In case the terms are separated and distributed over the source video, one fraction of the Intro part is created for the topic (here the Prime Minister), which covers half of T_{clip} . Then the remaining half is applied on the first identified term that follows the topic in the positive or negative term list of the event. For example, if the event is “The Prime Minister visits a foreign country” and the first matching term is “China”, this is the part on which the remaining shot time is applied to. If there are more than one section that can be chosen, the rule will apply that faces are more important than non-face videos (based on the fact that people recognize faces easier than landscapes). Note that although face detection has recently become a commodity technology, it was performed manually in the examples introduced

later in this paper.

• Final part

The clipping of the Final part basically follows the above rules. However, the Final part will always be the last to be edited as its duration is usually longer than that of other clips. The reason is that the first part of the negative event set, which asks for a controversy between Prime Minister and opposition, asks for the presentation of all opposition parties. Depending on the number of parties in Parliament, this might exceed the given temporal constraint, as clips need to be long enough to identify the different representatives of the parties (note that it is not necessary to cut on what they say, as the final soundtrack will be done manually).

◦ Topic

The generation of the Final part therefore begins with a recalculation of the actually used time, as some of the already trimmed parts could be shorter than the assigned time constraint. In that case, this remaining time is used for the Final part. The first part to be generated for the Final part is the actual topic, in our example, the Prime Minister’s resignation. For that, the terms “resignation”, “Prime Minister”, and “leaving office” are searched from the closed-captions. As all three terms are associated to a joint fraction in the story source, the “leaving office” part is trimmed backwards into the “Prime Minister’s speech”. If the “leaving office” part is too long, the window is shifted so long into the “Prime Minister’s speech” that the latter covers half of T_{clip} . If both parts are temporarily separated, they are individually located and half of the duration T_{clip} is assigned to each. They are then joined into one sequence, with the part of the Prime Minister resigning first.

◦ Controversy

Once the topic is covered, the remaining time is provided for trimming the controversy section. The story video is first searched for a sequence where the terms “Prime Minister” and “may resign” are detected. In case both appear around the same temporal proximity, the first appearance of each is taken as a start point. Then the occurrence of the term “opposition” or the different party names are searched for. Occurrences are marked, and the number of identified spots is summed up. One occurrence of a shot with the Prime Minister is added and the remaining time is devised by that figure, to allow for equal presentation. The trimming of each identified part starts at the end of each shot and extends in reverse direction for the calculated time. Only

Table 1 Values of common parameters for generating the examples.

Parameter	Value	Parameter	Value
$D_{\text{vid_length}}$	5 minutes	θ_s	2 stories
θ_{v_p}	60 %	P_{inaug}	2 time frames
θ_{c_p}	10 %	θ_{d_d}	7 days
θ_{v_n}	30 %	θ_{d_p}	80 %
θ_{c_n}	10 %	T_{poll}	3 seconds

the part of the Prime Minister follows the identification and overlay pattern, as the identified point may be part of a longer section. All the parts are then concatenated.

The final outcome is an editing list of video IDs and associated start and end times, as well as the IDs of inserts and their duration. The video generated with this framework at least needs an Intro part, one event in the middle and the Final part, all connected by two poll images. Thus, the shortest video possible that can be generated with this approach is 132 seconds long, resulting in the restriction of a journalist to not make use of this approach if the video should be even shorter. The generated video is the basis the journalist then starts working with, meaning that other events can be added, or the order or the material itself can be altered.

4. Examples

We applied the framework to three cases of resignation of Japanese Prime Ministers within the period from September 2006 to June 2009.

The actual processing time excluding the manual editing part for one Prime Minister was approximately 1 minute on a SUN Blade 100 work station. However, since most of this process could be done offline on the archive side at the time of capturing the video as a daily batch process, the proposed framework does not require much online processing.

Below, we introduce the three examples. For all the examples, the parameters were set as listed in **Table 1**.

4.1 Example 1: The case of Prime Minister Shinzo Abe

Fig. 6 is an example of a negative summary on Japanese Prime Minister Shinzo Abe.

The patterns searched for in the poll data are established based on an evaluation of the available poll data in combination with existing story fragments, resulting in the values in **Table 2**. The portrayed events represent explanations of pointed out patterns in the poll diagrams.

The actual events for example 1 are described in **Table 3**. Clips Te1-1 and Te1-2 are examples of the Intro part trimming, where clip Te1-1 represents the inaugu-

Table 2 Values of specific parameters for generating example 1.

Parameter	Value	Parameter	Value
N_{PM}	“Shinzo Abe”	P_{opening}	2 time frames
D_{inaug}	Sept. 26, 2006	N_{opening}	1 story
D_{resig}	Sept. 26, 2007	P_{closing}	3 time frames
N_{poll}	4	N_{closing}	2 stories
P	13 time frames	T_{clip}	24 seconds

Table 3 Events composing the summary in example 1. “To” events represent clips generated by the topic-based approach and “Te” events represent clips generated by the template-based approach.

Event	Description
Te1-1	Inauguration
Te1-2	Trip to China
To1-1	Return of party members after they had been expelled
To1-2	Meeting between party members to discuss the procedure of that return
To2-1	Prime Minister handles a crisis about a North Korean nuclear test
To2-2	Minister has to step down because he referred to women as “child-bearing machines”
To3-1	Government confirmed that pension records are mis-managed
To3-2	Crisis about North Korean refugees
To4-1	Opposition wins the election to the Upper House
Te2-1	US Airbase extension under discussion (this is the edited conflict sequence)
Te3-1	Prime Minister’s resignation (this is the resignation and leave office section)

ration speech and clip Te1-2 additional material, here the example of a trip to China. Clips To1-1 to To3-1 follow the general editing pattern explained in 3.4(1). Clip To3-2 contains a reverse ordering also described in 3.4(1). Clip To4-1 is again normally trimmed following the explanatory pattern. The keyframe for clip Te2-1 only shows the first part of an controversial topic sequence described in 3.5. The actual sequence of the opposition contains five clips, one for each opposition leader. Clip Te3-1 is another example of this part, as it contains the announcement of the Prime Minister about his resignation and a shot that shows him leaving the Parliament building.

We evaluated the video by comparing its content with the content and style of an explanation video actually created by NHK, the broadcaster (we call those further on “NHK video”). In the comparison with the NHK video about Prime Minister Abe’s resignation, where only two thirds of this version was accessible due to recording error, we found that clips Te1-1 to To1-2 covered the same events, though with different video material. Clip To2-1 was not at all used in the NHK video but instead a story about the resignation of a minister due to fraud. Clips To2-2 and To3-1 are then again covering the same events but with different video material.

Fig. 6 Example of a negative summary created by the proposed framework (N_{PM} = “Shinzo Abe”, (D_{inaug} , D_{resig}) = (Sept. 26, 2006, Sept. 26, 2007), D_{vid_length} = 5 minutes). Each clip is represented by a thumbnail image. Below are shown the clip ID, date of broadcast, and the story ID.

Table 4 Values of specific parameters for generating example 2.

Parameter	Value	Parameter	Value
N_{PM}	“Yasuo Fukuda”	$P_{opening}$	3 time frames
D_{inaug}	Sept. 26, 2007	$N_{opening}$	2 stories
D_{resig}	Sept. 24, 2008	$P_{closing}$	2 time frames
N_{poll}	3	$N_{closing}$	1 story
P	13 time frames	T_{clip}	29 seconds

Table 5 Events composing the summary in example 2. “To” events represent clips generated by the topic-based approach and “Te” events represent clips generated by the template-based approach.

Event	Description
Te1-1	Inauguration
Te1-2	Inauguration
Te1-3	Inauguration
To1-1	Lawsuit against government on hepatitis infection by contaminated blood products
To2-1	Discussion about the problem of a special budget on roads
To2-2	Enactment of a tax bill
Te2-1	Resignation
Te2-2	Resignation (comments of the opposition)
Te2-3	Resignation (comments of the general public)

The event of clip To3-2 in our video is again not used in the NHK video, which addresses instead the suicide of a cabinet Minister. For the period of our clips To4-1 to Te3-1, unfortunately, there was no comparison available. It is also apparent that the NHK uses less strict pattern of insertion of demoscopic material. The pattern in the NHK video was three events, and then a succession of two events separated by poll inserts. An additional difference was that the inserts in the NHK video were animated, though the presented information was static. This difference is also applicable for all other videos too.

The generated video was less complex compared to the related NHK video, with respect to selected events as well as clips.

4.2 Example 2: The case of Prime Minister Yasuo Fukuda

Fig. 7 is an example of a negative summary on Japanese Prime Minister Yasuo Fukuda.

The patterns searched for in the poll data are established based on an evaluation of the available poll data in combination with existing story fragments, resulting in the values in **Table 4**. The portrayed events represent explanations of pointed out patterns in the poll diagrams.

This second generated video about Prime Minister Fukuda demonstrates the flexibility of the template approach for the overall generation of the video. This clip is relevant, as the amount of critical poll events

supposed to be two based on our selection process (see steps 2, 3 and 4 in 3.4 (1)), but only one relevant event for the first explanation could be found. Due to this mismatch of values between assigned and actually retrieved events, the exception-rule fires that allows altering the number of clips for the Intro and Final Part during later editing, so that the set clip time constraint can be kept.

Thus, for the Intro, three sequences were chosen for the inauguration. They were all cut from the same sequence, one from the beginning, one from the middle and one from the end.

In the Final part, we see an interesting effect of the rule and template system. First the resignation sequence is built according to the template. Now, as the resignation is the crisis that needs to be commented on, the rule will fire that the comments of the opposition needs to be added. However, there is still time left after that. Thus, an additional sequence is added, which, according to the template, represents the opinion of the general public. Even though the end result provides an individualized form for this particular case, it needs to be said that the final video is too long and repetitive, hence aesthetically not pleasing. However, as there are no rules to override the given constraints about the length, this video demonstrates that the template approach facilitates a sensible finalization of the video, in particular in situations where the main material is not optimally distributed.

Overall, this video covers the same events as the NHK counterpart, except the public sequence at the end, which was not part of the NHK video, only that again different material was chosen. The NHK video was much better trimmed, which shows that our algorithm is still to stringent in its application of rhythm and user expectations with respect to shot length.

4.3 Example 3: The case of Prime Minister Taro Asou

Fig. 8 is an example of a negative summary on Japanese Prime Minister Taro Asou.

The patterns searched for in the poll data are established based on an evaluation of the available poll data in combination with existing story fragments, resulting in the values in **Table 6**. The portrayed events represent explanations of pointed out patterns in the poll diagrams. The actual events for example 3 are described in **Table 7**.

In this third video about Prime Minister Asou, here

Fig. 7 Example of a negative summary created by the proposed framework ($N_{PM} = \text{"Yasuo Fukuda"}$, $(D_{inaug}, D_{resig}) = (\text{Sept. 26, 2007, Sept. 24, 2008})$, $D_{vid_length} = 5$ minutes). Each clip is represented by a thumbnail image. Below are shown the clip ID, date of broadcast, and the story ID.

again the template approach reacts flexible to the generation of the Intro, which covers the inauguration speech only. The generation of the three different explanation events does show the expected pattern, namely two events per explanation. It needs to be stated, though, that the third event represents a positive event, namely a number of foreign journeys to important meetings. As

this period of government was, similar to Prime Minister Fukuda's period, rather short, the final representation of the main crisis event and the final resignation are merged into one block. Note that in this version, no material of the opposition is included, because due to the circumstances, no statements were available. Again, the collected events are the same as in the NHK video,

Table 6 Values of specific parameters for generating example 3.

Parameter	Value	Parameter	Value
N_{PM}	“Taro Asou”	$P_{opening}$	3 time frames
D_{inaug}	Sept. 24, 2008	$N_{opening}$	2 stories
D_{resig}	Sept. 16, 2009	$P_{closing}$	5 time frames
N_{poll}	3	$N_{closing}$	3 stories
P	14 time frames	T_{clip}	24 seconds

Table 7 Events composing the summary in example 3. “To” events represent clips generated by the topic-based approach and “Te” events represent clips generated by the template-based approach.

Event	Description
Te1-1	Inauguration
Te1-2	Inauguration
To1-1	Comments that some medical doctors are socially problematic
To1-2	First debate with opposing party leader on additional budget plan
To2-1	Attends the Summit on Financial Markets and the World Economy in London
To2-2	Plans to dismiss the Parliament
To3-1	North Korea conducts an underground nuclear test
To3-2	Expresses concern at the UN Security Council
Te2-1	Discusses about occupied islands with Russian President in Italy
Te2-2	Announces to dismiss the Parliament
Te2-3	Resigns the leader of the party after the loss in the election

which can be explained that the time period of the government is too short to pose the option of choice to the system. This also underpins the fact that the chosen thresholds and constraints in our design make sense.

We also asked Japanese colleagues to state how reasonable the created videos were (we showed the videos and provided them with a comment about the event portrayed in each of the videos). All stated that the presented events could explain the pattern indicated in the preceding poll but if the events were always the most important they did not dare to say. It would all depend on the voice-over, as this finally provides the contextualization of the event in a larger story.

This type of evaluation is, of course, rather rudimentary. A more fine grained qualitative and quantitative evaluation procedure is currently under development. The aim is to investigate with professionals, i.e. journalists and editors, as well as with potential viewers of the clips, how the various levels of perception, i.e. content and style, are graded by the different type of users. This is not an easy endeavor, as the relevance of events in a domain like politics as well as presentation style, necessarily have a layer of subjectivity in it. Yet, the evaluation we currently work towards provides the participants with all available resources to facilitate an analysis of the approach that covers measurable as-

pects, such as preferred length of a video and its visual representation, or the relevance of events, in combination with explanatory open ended questions that should help us to better capture the elements of human reasoning behind the selection and juxtaposition of news material within the context of story telling. These tests are currently under way.

5. Discussion

The approach described in this paper is but a small step towards the intelligent editing of explanatory news video. Nevertheless, we believe that the work undertaken informs research into the automated generation of video beyond the summarization of a single news show. The proposed framework in the current state describes the general pipeline of steps to generate event explanations for the domain of news in a chronological and retrospective manner. As the approach works in real-time, the system takes the first established story. We demonstrated that the coupling of topic-based and template-based automatic authoring works well for a video repository that contains already edited material for a consistent domain, as in our case, news. The chronological approach of representing events in combination with story threads can also be applied to similar description structures, such as the review of a year (What were the highlights of 2012?) or the collection of major events in a person’s life (Why was Elisabeth Taylor famous?). In particular the latter example is what we aim for next, namely larger story structures in form of obituaries or tributes.

The achieved results with respect to event identification and automatic retrieval of related material propose that the framework can be utilized to ease the work of journalists, i.e. leave it to the system to find the essential events for a particular case of explanation and provide this as a start point for the journalist to then finalize the video. The three examples of generated material are, with respect to structure and content, comparable to those versions generated by professionals at a broadcasting company. Besides providing a larger set of templates and technology to establish those, it would be interesting to see if professionals (editors or journalists) would save time with the proposed framework, assuming that some re-editing and the occasional search for better material might occur. This question, however, can only be answered with a long-time user test, which is beyond the current state of the environment.

With the current state of the framework, we could

Fig. 8 Example of a negative summary created by the proposed framework (N_{PM} = “Taro ASOU”, (D_{inaug}, D_{resig}) = (Sept. 24, 2008, Sept. 16, 2009), D_{vid_length} = 5 minutes). Each clip is represented by a thumbnail image. Below are shown the clip ID, date of broadcast, and the story ID.

demonstrate that the template-approach allows a directed access to relevant data sources that then facilitates the development of the explanatory video. In the presented case, the demographic poll data allowed the identification of significant points in time that potentially can indicate pointers to relevant events. It needs to be seen if other topics in politics in particular or other domains like entertainment provide similar

semantic correlations between the topic and a presentation form that facilitates a temporal as well as semantic structure to establish an explanation for the topic. It needs to be stated that the system will not work in case the demographic data is not available. However, different sources can provide similar data, e.g. demographic analysis of Twitter tweets. It then needs to be investigated how these sources can take the role of triggers

for the related events.

As also demonstrated in this paper, our approach might not necessarily generate the appropriate events and might also establish a different presentation form (i.e. using different video material) as those chosen by professional editors or journalists. Though the response on our generated examples showed that the events made sense, we are also aware of the fact that the news domain is a thankful medium, where the visual content merely provides a carpet for the dominant voice-over. The relation between image and audio, therefore, needs further investigation. At the moment, all we can state is that the framework introduces a biased angle, though that is again referring to the closed-caption and hence to the soundtrack. Additional descriptors need to be introduced into the template that can support a similar sentiment analysis as is used for the transcript evaluation.

We need to admit that the formalization of the current template schema is rather ad-hoc. The design is mainly based on the analysis of the particular event-type we were interested in. Even though the overall approach of using contextualized keywords works, it needs to be investigated how far this can be correlated to the ontologies already in use within broadcasting organizations. In the same context, it is not yet clear if the strict planning of a fixed presentation pattern does not result in neglecting significant material. The overall format, though, is appropriate as it reflects the common broadcasting style for presenting such types of explanations. It is worthwhile to investigate how different presentation forms would influence the structure of the templates associated to a topic. In this context, it would also be interesting to investigate how the already existing audio track can be used for the selection of video clips or for the final presentation generation. In particular, the latter can contribute to an approach towards complete automation of the production process.

Though the described approach works, we still need to verify with journalists in how much it fits their needs. In particular, the ways how they could alter the structures easily if they wish to work with different events needs to be investigated. We currently consider to either automatically provide, or facilitate the user to edit, alternative storylines.

6. Conclusion

We presented a framework for the semi-automated authoring of explanatory audio-visual news topics in a

retrospective style for the domain of politics, which follows a hybrid approach that couples an automatic topic-based approach of story generation with template-based generation of the introduction and the final parts of the video. We demonstrated that the approach works for in an environment that can depend on already edited material, such as the news repository of a broadcaster. The next further developments will focus on the integration of the developed technologies into a complete system and further research on additional template types.

7. Acknowledgments

Parts of this work were supported by the Grants-in-Aid for Scientific Research (21013022, 21700103, 2300036), the Excellent Young Researchers Overseas Visit Program (21-5324) from JSPS, and a joint research project between Nagoya University, University of Amsterdam, NetCompass, Ltd. and National Institute of Informatics.

References

- 1) P. Appan and H. Sundaram : "Networked Multimedia Event Exploration", *Proc. 12th ACM International Conference on Multimedia*, pp.40–47 (Oct. 2004)
- 2) J. Bateman, J. Kleinz, T. Kamps, and K. Reichenberger : "Towards Constructive Text, Diagram, and Layout Generation for Information Presentation", *Comput. Linguist.*, 27, pp.409–449 (2001)
- 3) S. Bocconi and F. Nack : "Vox Populi: Automatic Generation of Biased Video Sequences", *Proc. ACM Multimedia 2004 Workshop of Story Representation, Mechanism and Context*, pp.9–16 (Oct. 2004)
- 4) D. C. A. Bulterman and L. Hardman : "Structured Multimedia Authoring", *ACM Trans. Multimedia Comput. Commun. Appl.*, 1, pp.89–109 (2005)
- 5) M. G. Christel, A. G. Hauptmann, H. D. Wactler, and T. D. Ng : "Collages as Dynamic Summaries for News Video", *Proc. 10th ACM International Conference on Multimedia*, pp.561–569 (Dec. 2002)
- 6) G. Davenport and M. Murtaugh : "ConText: Towards the Evolving Documentary", *Proc. 3rd ACM International Conference on Multimedia*, pp.381–389 (Nov. 1995)
- 7) E. Dumont, B. Mérialdo, S. Essid, W. Bailer, H. Rehatschek, D. Byrne, H. Bredin, N. E. O'Connor, G. J. F. Jones, A. F. Smeaton, M. Haller, A. Krutz, T. Sikora, and T. Piatrik : "Rushes Video Summarization Using a Collaborative Approach", *Proc. 2nd ACM TRECVID Video Summarization Workshop*, pp.90–94 (Oct. 2008)
- 8) J. G. Fiscus and G. R. Doddington : "Topic Detection and Tracking Evaluation Overview", J. Allan, ed., *Topic Detection and Tracking*, Kluwer Academic Publishers, Norwell, MA, USA, pp.17–31 (2002)
- 9) J. R. Gregory : "Some Psychological Aspects of Motion Picture Montage", Ph.D. Thesis, University of Illinois (1961)
- 10) L. Hardman, Z. Obrenović, F. Nack, B. Kerhervé, and K. Pierisol : "Canonical Processes of Semantically Annotated Media Production", *Multimedia Systems*, 14, pp.327–340 (2008)
- 11) C. Hargood, D. E. Millard, and M. J. Weal : "A Semiotic Approach for the Generation of Themed Photo Narratives", *Proc. 21st ACM Conference on Hypertext and Hypermedia*, pp.19–28 (June 2010)
- 12) A. G. Hauptmann, R. Yan, and W.-H. Lin : "How Many High-level Concepts will Fill the Semantic Gap in News Video Retrieval?", *Proc. 6th ACM International Conference on Image and Video Retrieval*, pp.627–634 (July 2007)
- 13) I. Ide, T. Kinoshita, T. Takahashi, H. Mo, N. Katayama,

- S. Satoh, and H. Murase : “Efficient Tracking of News Topics based on Chronological Semantic Structures in a Large-scale News Video Archive”, *IEICE Trans. on Information and Systems*, E95-D, 5, pp.1288–1300 (2012)
- 14) N. Katayama, H. Mo, I. Ide, and S. Satoh : “Mining Large-scale Broadcast Video Archives towards Inter-video Structuring”, *Advances in Multimedia Information Processing —PCM2004 5th Pacific Rim Conference on Multimedia Proceedings, Part II, Lecture Notes in Computer Science*, 3332, pp.489–496 (Dec. 2004)
 - 15) F. Nack : “AUTEUR: The Application of Video Semantics and Theme Representation for Automated Film Editing”, Ph.D. Thesis, Lancaster University (Aug. 1996)
 - 16) F. Nack and W. Putz : “Designing Annotation before It’s Needed”, *Proc. 9th ACM International Conference on Multimedia*, pp.251–260 (Oct. 2001)
 - 17) F. Nack and I. Ide : “Why did the Prime Minister Resign? —Generation of Event Explanation from Large News Repositories—”, *Proc. 19th ACM International Conference on Multimedia*, pp.313–322 (Nov. 2011)
 - 18) S.-Y. Neo, Y. Ran, H.-K. Goh, Y. Zheng, T.-S. Chua, and J. Li : “The Use of Topic Evolution to Help Users Browse and Find Answers in News Video Corpus”, *Proc. 15th ACM International Conference on Multimedia*, pp.198–207 (Sept. 2007)
 - 19) M. O. Riedl and R. M. Young : “From Linear Story Generation to Branching Story Graphs”, *IEEE Computer Graphics and Applications*, 26, 3, pp.23–31 (2006)
 - 20) L. Rutledge, M. Alberink, R. Brussee, S. Pokraev, W. van Dieten, and M. Veenstra : “Finding the Story: Broader Applicability of Semantics and Discourse for Hypermedia Generation”, *Proc. 14th ACM Conference on Hypertext and Hypermedia*, pp.67–76 (Aug. 2003)
 - 21) A. Scherp and S. Boll : “Context-driven Smart Authoring of Multimedia Content with xSMART”, *Proc. 13th ACM International Conference on Multimedia*, pp.802–803 (Nov. 2005)
 - 22) A. Scherp, T. Franz, C. Saathoff, and S. Staab : “F—a Model of Events Based on the Foundational Ontology dolce+DnS Ultra-light”, *Proc. 5th International Conference on Knowledge Capture*, pp.137–144 (Sept. 2009)
 - 23) A. Scherp, R. Jain, M. Kankanhalli, and V. Mezaris : “Modeling, Detecting, and Processing Events in Multimedia”, *Proc. 18th ACM International Conference on Multimedia*, pp.1739–1740 (Oct. 2010)
 - 24) F. Shipman, A. Girgensohn, and L. Wilcox : “Generation of Interactive Multi-level Video Summaries”, *Proc. 11th ACM International Conference on Multimedia*, pp.392–401 (Nov. 2003)
 - 25) C. G. M. Snoek, M. Worring, and A. G. Hauptmann : “Learning Rich Semantics from News Video Archives by Style Analysis”, *ACM Trans. Multimedia Comput. Commun. Appl.*, 2, pp.91–108 (2006)
 - 26) H. Takamura, T. Inui, and M. Okumura : “Extracting Semantic Orientations of Words Using Spin Model”, *Proc. 43rd Annual Meeting of the Association for Computational Linguistics*, pp.133–140 (June 2005)
 - 27) M. Theune, S. Faas, A. Nijholt, and D. Heylen : “The Virtual Storyteller: Story Creation by Intelligent Agents”, *Proc. 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment*, pp.204–215 (Mar. 2003)
 - 28) S. Toulmin, R. Rieke, and A. Janik : “An Introduction to Reasoning”, MacMillan Publishing Company, 2nd ed. (1984)
 - 29) U. Westermann, S. Agaram, B. Gong, and R. Jain : “Event-centric Multimedia Data Management for Reconnaissance Mission Analysis and Reporting”, *Proc. 14th ACM International Conference on Multimedia*, pp.499–500 (Oct. 2006)

Ichiro Ide Dr. Ichiro Ide received his B.Eng., M.Eng., and PhD from The University of Tokyo in 1994, 1996, and 2000, respectively. He became an Assistant Professor at the National Institute of Informatics, Japan in 2000. Since 2004, he has been an Associate Professor at Nagoya University. He had also been a Visiting Associate Professor at National Institute of Informatics from 2004 to 2010, an Invited Professor at Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), France in 2005, 2006, and 2007, a Senior Visiting Researcher at ISLA, Instituut voor Informatica, Universiteit van Amsterdam from 2010 to 2011. His research interest ranges from the analysis and indexing to retargeting of multimedia contents, especially in large-scale broadcast video archives, mostly on news, cooking, and sports contents. He has been serving on program committees at conferences such as ACM MM, ICME, CVIM, MMM, and ICMR. He is a senior member of IEICE, and members of IPS Japan, JSAI, ANLP, IEEE, and ACM.

Frank Nack Dr. Frank Nack is tenure assistant professor at the Intelligent System Lab Amsterdam (ISLA) of the Informatics Institute at the University of Amsterdam (UvA). He obtained his Ph.D. from Lancaster University, UK in 1996. The main thrust of his research is on context and process aware media knowledge spaces (design, interaction, memory); the representation and adaptation of experiences in mobile and ambient environments; hypermedia systems that enhance human communication and creativity; interactive storytelling; computational applications of media theory and semiotics; and computational humour theory. He published more than 100 papers on these topics. He has been serving on program committees (member or chair) at the major conferences: ACM MM, ACM Hypertext, Multimedia Modeling (MMM), ICME, and ICIDS. He frequently reviews for ACM TOMCCAP, Multimedia Tools and Applications, The Information Retrieval Journal, IEEE Intelligent Systems Magazine, IEEE Multimedia, International Journal of Continuing Engineering Education and Life Long Learning, New Review of Hypermedia and Multimedia. He is on the editorial board of IEEE Multimedia, where he edits the Media Beat column and also serves as associated editor in chief. He is member of the ACM, ACM SIGMM, ACM SIGCHI and ACM SIGWEB.