Why Did the Prime Minister Resign?

Generation of Event Explanations from Large News Repositories

Frank Nack  
ISLA, Informatics Institute  
University of Amsterdam  
Postbus 94323,  
1098 XH Amsterdam, The Netherlands  
nack@uva.nl

Ichiro Ide∗  
Graduate School of Information Science  
Nagoya University  
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan  
ide@is.nagoya-u.ac.jp

ABSTRACT

One of the common parts of news is to provide the background for a current event, such as the resignation of a Prime Minister. This paper addresses a framework that facilitates semi-automated authoring of explanatory audio-visual news topics in a retrospective style for the domain of politics based on already edited new stories available in the repository of the news corporation. The aim is to facilitate a journalist with an audio-visual body based on which he/she can finalize the explanatory piece. The proposed framework enhances current state of the art video summarization by allowing the combination of different news stories into one coherent explanation about a topic of the current news. 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.

Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]: Video; I.7.2 [Document Preparation]: Multi/mixed media

General Terms

Human Factors, Design

Keywords

Multimedia authoring, template, topic, topic thread, event, automated video editing, story generation, video archive

∗This work was done during his stay at Informatics Institute, University of Amsterdam.

∗Area chair: Susanne Boll

1. INTRODUCTION

A news program on TV is regularly scheduled and reports current events. They follow typically 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. A news program can include live material, e.g. recorded interviews by field reporters, expert opinions, and opinion polls, or editorial content, i.e. short background information that explains a current event.

In particular, the editorial content for short background information in form of audio-visual material requires a complex and resource demanding production process, structured for very rapid assembly of material from very diverse sources [13]. The aim is to facilitate background information to a headline of the news program in form of an explanatory news story, for example about the resignation of a national government. This type of story needs to explain within a given time span 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 clip 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 story represents 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 processes [8]. 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 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 the basic elements of the framework. 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 material through automated editing. Finally, we assess our achievements thus far, 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 retrieval [9, 19], which is relevant for the detection of topics in non-annotated material. Other works cover the aspect of video summarization [4, 5, 18], which address the problem of distilling appropriate videos from large corpora. Work on topic clustering points into the same direction [14], 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" [10] 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 connecting different news stories into an explanatory video for a particular topic currently discussed in the news.

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 [3, 8, 12, 13, 15]. All these works address the various steps within multimedia production [8], always assuming that there is an author involved in the process. We also assume an author in the production loop, in particular for the making of the voice over, but 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) [8].

In recent years, substantial work has been provided for the modeling of multimedia events [1, 17, 21]. 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 (see in particular [16]). 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-automated 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 NIST [6] and the work by Ide et al. [10]. 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 topicically 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 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. Figure 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 video.

For the ease of readability, we will use the resignation of a Prime Minister 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 negative. However, a positive video could be generated by reverting the bias to positive and then apply the related negative. However, a positive video could be generated by reverting the bias to positive and then apply the related negative.

### 3.2 Dataset

The video data used in our work comes from the NII-TVRECS archive [11], which consists of a Japanese news show recorded daily for more than ten years (1,800 hours). The video data is accompanied with closed-caption, and is segmented into stories by the method described in [10].

From this archive, the proposed framework extracts all texts during the period. As shown in Fig. 1, a topic thread structure represents the development of news topics along the time-line concerning a story under consideration.

### 3.3 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 demoscopic polls (see an example in Fig. 3) as an external reference that should establish the key moments to be considered. As the presented opinions relate to real events, we try to identify those and in that way we make use of the collective memory of the viewers, as the used clips function as “triggers” to well-known events during the time period of the topic that needs to be explained. The duration of the polls follows the time-frame $p_i$ (for $i = 1, 2, ..., P$) with a period of approximately one month between two consecutive monthly poll measurements while the 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.

#### 3.4.1 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:

![Figure 3: Example of the transition of polls ($N_{PM} = \text{“Shinzo Abe”}$).](image)

1. A Japanese morphological analyzer JUMAN developed at Kyoto University was used. It is available for download from [http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman-e.html](http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman-e.html).
2. 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](http://nlp.kuee.kyoto-u.ac.jp/nl-resource/knp-e.html).
1. Time-frames \( p_i \ (i = 1, ..., P) \), 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 represent the time period.

Details of the four steps follows.

**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\(^3\). The polls consist of ratios on the approval (“Yes”) or the disapproval (“No”) of the government led by the Prime Minister in focus. Figure 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 pattern**
  - The “Yes” rate rises above \( \theta_{vp} \).
  - The “Yes” rate increases more than \( \theta_{vp} \) per time-frame.
  - The “Yes” rate exceeds the “No” rate.

- **Negative poll pattern**
  - The “Yes” rate drops below \( \theta_{vn} \).
  - The “Yes” rate decreases more than \( \theta_{vn} \) 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.

**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 talk about 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 alternation between poll graph description and explanatory clips is then followed in alternating fashion until the end, where the video concludes 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 and the cognitive complexity of the potential clips (e.g. the image of a poll graph can be comprehended within \( T_{poll} = 3 \) seconds, where an event clip is usually composed of shots with different frame ratios and varying content complexity resulting in a length of somewhere between 30 to 50 seconds \([7]\)).

The calculation in our framework is as such: The start point is the overall time of the final video \( D_{vid\_length} \). From that we subtract \( N_{poll} T_{poll} \), and the remaining time is divided into equal blocks covering the introduction, the explanatory events, and the conclusion, where we follow Gregory’s setup of shot ratio comprehension \([7]\). 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.

**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. \([10]\). 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.

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:

---

3 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/yoron/seijiishiki/list_seijiishiki1.html](http://www.nhk.or.jp/bunken/research/yoron/seijiishiki/list_seijiishiki1.html), and also published in “The NHK Monthly Report on Broadcast Research”.

---

![Figure 4: Terminal/original story in a topic thread structure.](image-url) Stories concerning general structures such as the one shown in Fig. 1 are neither a terminal nor an original 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.

**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 words\(^4\) [20] 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 $\theta_s$ stories for a positive poll pattern.
- Select the least $\theta_s$ stories for a negative poll pattern.

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

The outcome of the presented four steps is an editing list that contains the IDs of the stories to be included in the final output video. 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.

### 3.4.2 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 be hence 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". Showing this is in fact a convention among broadcasters for news related explanatory videos, so much so that a particular term is available to make this sequence easily retrievable.

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 this (the terms in [ ] describe some of the concepts we search for):

- Inauguration speech (this is always shown) or if not available use the Prime Minister talking in Parliament [Prime Minister, \text{NP\_PM}, Inauguration], alternative [Prime Minister, \text{NP\_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, \text{NP\_PM}, China, US, Germany, ...]
  - The Prime Minister announces a positive budget [Prime Minister, \text{NP\_PM}, budget, positive]
  - The Prime Minister presents a positive economical view [Prime Minister, \text{NP\_PM}, good economy, improving economy, industry, presentation]
  - The Prime Minister wins a prize [Prime Minister, \text{NP\_PM}, prize, receive, ceremony]
  - Government party wins a by-election [Prime Minister, \text{NP\_PM}, election, win]

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

- Conflict with the opposition and Prime Minister resigns [Prime Minister, \text{NP\_PM}, may resign, opposition party names, leader, party, criticize] and [Prime Minister, \text{NP\_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, \text{NP\_PM}, health, hospital, way to hospital]
  - Government Party loses a by-election [Prime Minister, \text{NP\_PM}, election, loss]
  - The Prime Minister or a minister has juridical problems [Prime Minister, \text{NP\_PM}, fraud, party members, ministers]
  - The Prime Minister announces a negative budget [Prime Minister, \text{NP\_PM}, budget, negative]
  - The Prime Minister presents a negative economical view [Prime Minister, \text{NP\_PM}, weak economy, high inflation, deflation, industry, presentation]

The amount of elements to be presented depends, in accordance to Step 2 in 3.4.1, on the temporal duration of non-covered 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 \( \{p_1, ..., p_{\text{inaug}}\} \), the first story that contains the term “inauguration speech” (Number of stories \( N_{\text{inaug}} = 1 \)).

- **Opening stories**
  Depending to the period \( \{p_1, ..., p_{\text{opening}}\} \) before the occurrence of the first positive/negative poll pattern, \( N_{\text{opening}} \) 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 \( \{p_{P-\text{closing}+1}, ..., p_P\} \) after the poll has dropped below \( \theta_{\text{pol}} \), \( N_{\text{closing}} \) 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.

#### Step 1: Generation of the graphical inserts.

The graphical inserts serve as the introduction of significant turning points and facilitate the corner stones of the retrospective structure. Figure 5 contains four different examples of such inserts. 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 in [2]. However, they are currently generated manually.

#### 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 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

\[
T_{\text{clip}} = \frac{T - N_{\text{poll}}T_{\text{poll}}}{\theta_{\theta}N_{\text{poll}} + N_{\text{inaug}} + N_{\text{opening}} + N_{\text{closing}}}. \tag{1}
\]

The automatic cropping in our framework follows the approach described in [13], which provides a planner for news composition based on 22 rules for the automated clipping of a shot and juxtaposition of shots. In [13], 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.

For the explanatory videos, we search in every clip for the topic (here the Prime Minister’s name \( N_{\text{PM}} \)) in the closed-caption. The related time code is taken and then a duration of \( T_{\text{clip}} \) 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 intro section of the anchorperson and take the first \( T_{\text{clip}} \) into that part of the story video. The assumption is that this is the main event to which the poll refers to.

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. 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 \( T_{\text{clip}} \) 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 \( T_{\text{clip}} \) 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 \( T_{\text{clip}} \). 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).
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 the 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 amount 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).

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 the identified part. If both parts are temporarily separated, they are individually located and half of the duration is assigned to each. They are then joined into one sequence, with the part of the Prime Minister resigning first.

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 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 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 does not require much online processing.

### Table 1: Values of parameters for generating the example(s).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{PM}$</td>
<td>“Shinzo ABE”</td>
</tr>
<tr>
<td>$D_{inaug}$</td>
<td>Sept. 26, 2006</td>
</tr>
<tr>
<td>$D_{resig}$</td>
<td>Sept. 26, 2007</td>
</tr>
<tr>
<td>$D_{vid,ength}$</td>
<td>5 minutes</td>
</tr>
<tr>
<td>$\theta_{sp}$</td>
<td>60 %</td>
</tr>
<tr>
<td>$\theta_{dp}$</td>
<td>80 %</td>
</tr>
<tr>
<td>$\theta_{sp}$</td>
<td>10 %</td>
</tr>
<tr>
<td>$\theta_{dp}$</td>
<td>24 seconds</td>
</tr>
<tr>
<td>$\theta_{dp}$</td>
<td>3 seconds</td>
</tr>
<tr>
<td>$\theta_{op}$</td>
<td>30 %</td>
</tr>
<tr>
<td>$\theta_{dp}$</td>
<td>1 story</td>
</tr>
<tr>
<td>$\theta_{op}$</td>
<td>30 %</td>
</tr>
<tr>
<td>$\theta_{dp}$</td>
<td>3 time frames</td>
</tr>
<tr>
<td>$\theta_{op}$</td>
<td>2 stories</td>
</tr>
<tr>
<td>$\theta_{dp}$</td>
<td>24 seconds</td>
</tr>
</tbody>
</table>

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

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

in Table 1. The portrayed events represent explanations of pointed out patterns in the poll diagrams.

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.

The actual events are described in Table 2. Clips To1-1 and To1-2 are examples of the Intro part trimming, where clip To1-1 represents the inauguration speech and clip To1-2 additional material, here the example of a trip to China. Clips T1-1 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 To2-1 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 To3-1 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.

### 4. EXAMPLE

We applied the framework to four cases of resignation of Japanese Prime Ministers within the period from September 2006 to June 2010. Figure 5 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

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 the 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 amount 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).

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 the identified part. If both parts are temporarily separated, they are individually located and half of the duration is assigned to each. They are then joined into one sequence, with the part of the Prime Minister resigning first.

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 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 is 132 seconds long. 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. EXAMPLE

We applied the framework to four cases of resignation of Japanese Prime Ministers within the period from September 2006 to June 2010. Figure 5 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

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 the 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 amount 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).

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 the identified part. If both parts are temporarily separated, they are individually located and half of the duration is assigned to each. They are then joined into one sequence, with the part of the Prime Minister resigning first.

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 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 is 132 seconds long. 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. EXAMPLE

We applied the framework to four cases of resignation of Japanese Prime Ministers within the period from September 2006 to June 2010. Figure 5 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
Figure 5: Example of a negative summary created by the proposed framework ($N_{PM} = \text{“Shinzo ABE”}$, $(D_{inaug}, D_{resig}) = (\text{Sept. 26, 2006, Sept. 26, 2007})$, $D_{vid\_length} = 5$ minutes). Each clip is represented by a thumbnail image. Below are shown the clip ID, date of broadcast, and the story ID.
We evaluated this clip and one other of the four generated explanatory videos by comparing their content with the content and style of explanation videos 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, 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. 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 the same. Both ours and the second NHK video were much shorter than the video of Fig. 5, as Prime Minister Kunio Hatoyama only showed dropping poll rates. The two videos showed hence the same pattern, the same events but again with different video material. We also asked Japanese colleagues to state how reasonable the created videos were (we showed the video and provided them with a comment about the event portrayed in the video). 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 all depended on the voice over.

5. DISCUSSION

The approach described in this article 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 1999?) 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.

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 two 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 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 demoscopic 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 demoscopic data is not available. However, different sources can provide similar data, e.g. demoscopic 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 contextualised 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, 23000036), 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.

8. REFERENCES


