Threading news video topics

Ichiro IDE
National Inst. of Informatics
2–1–2 Hitotsubashi,
Chiyoda-ku, Tokyo 101–8430,
Japan
ide@nii.ac.jp

Hiroshi MO
National Inst. of Informatics
2–1–2 Hitotsubashi,
Chiyoda-ku, Tokyo 101–8430,
Japan
mo@nii.ac.jp

Norio KATAYAMA
National Inst. of Informatics
2–1–2 Hitotsubashi,
Chiyoda-ku, Tokyo 101–8430,
Japan
katayama@nii.ac.jp

ABSTRACT

We introduce a method to extract topic threads throughout a large-scale news video corpus as well as an interface that provides the users with the facility to browse through the corpus guided by the thread structure. The thread-based interface enables thorough understanding of a topic of interest, and moreover, it is designed to reduce the number of candidates provided during the tracking process which lightens the burden of selection imposed on the users. In this paper, we introduce the details of topic segmentation, tracking/threading, followed by the evaluation of the segmentation and the introduction of a prototype interface. The evaluation showed practical ability, and the trial use of the interface turned out to be effective and informative.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*data mining*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*information filtering, search process, selection process*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*text analysis*

General Terms

Design, Experimentation, Algorithms

Keywords

Multimedia information retrieval, video corpus, news video, topic segmentation, topic tracking, topic threading, user interface

1. INTRODUCTION

Following recent advance in telecommunication technology, large amount of video data has become available through various media. Such video data contain a broad range of human activities, which could be considered as valuable cultural and social heritage of the human race. News videos contain such information most densely, but inter-video semantic analysis throughout a large-scale news video corpus has not been examined in majority of previous works.

We have built an automatic news video archiving system where important topics can be searched and tracked easily. It automatically records and archives video image, audio, and closed-caption text transmit from broadcasting stations. Up to now, approximately 410 hours (254GB of MPEG-1 and 1.55TB of MPEG-2 format videos, and 18.7MB of closed-caption texts) have been obtained and archived from a specific Japanese daily news program.

In this paper, we introduce a method to structure a news video corpus by topic threading as well as a user interface that provides the users with the ability to browse through the corpus based on the structure. The interface not only enables thorough understanding of the topic of interest, but also reduces the number of selections imposed on the users, which is essential for effective browsing through a large-scale video corpus.

2. STRUCTURING A NEWS VIDEO CORPUS BY TOPIC THREADING

2.1 Structure of news video corpus

A news video corpus may seem merely as an accumulation of daily news videos. Since this appearance is natural when handling a small set of videos, majority of previous works concentrated on video segmentation (*i.e.* intra-video structuring) to obtain minimal semantic units (stories). However, when the size of the corpus expands, it is no longer a mere accumulation of independent stories, but a group of stories mutually related among themselves, where the relation itself provides rich semantic information. Figure 1 shows an example of the two structures in the corpus. The left half shows the intra-video structure, and the right half shows the inter-video structure throughout the corpus based on the relations between the intra-video structure.

In this paper, we propose a method for inter-video structuring, namely topic threading, as well as a user interface based on the thread structure that provides the users with the ability to interactively track up and down a topic of interest by selecting topic threads. Figure 2 shows an example of preceding and succeeding topic threads originating from a story (In this case, the first topic on March 1, 2002), actually obtained from the corpus.

MIR'03, November 7, 2003, Berkeley, California, USA.

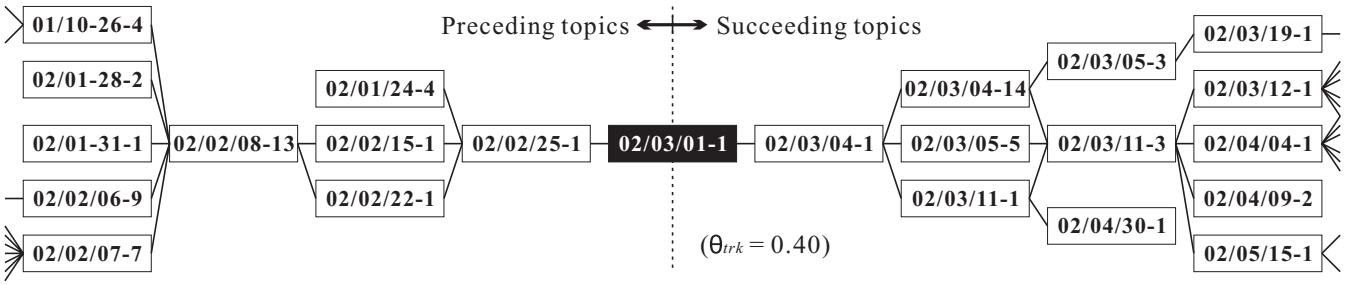

Figure 2: Part of the topic thread structure extracted from the corpus. Topics are named in the following format "Year/Month/Day-Topic#". Summaries of actual contents are shown in detail in Figure 10.

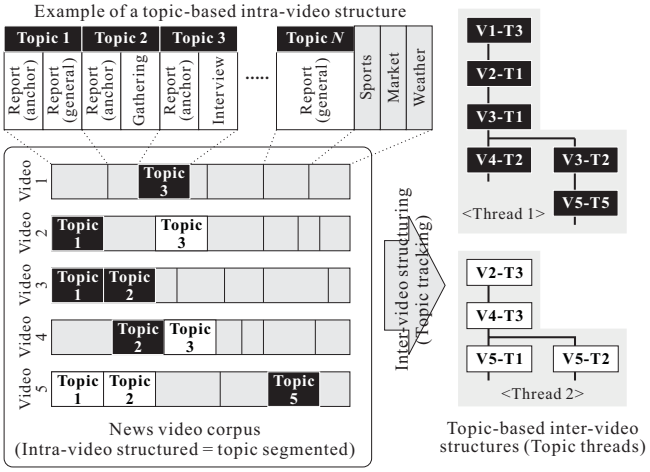

Figure 1: Topic-based intra-/inter-video structures.

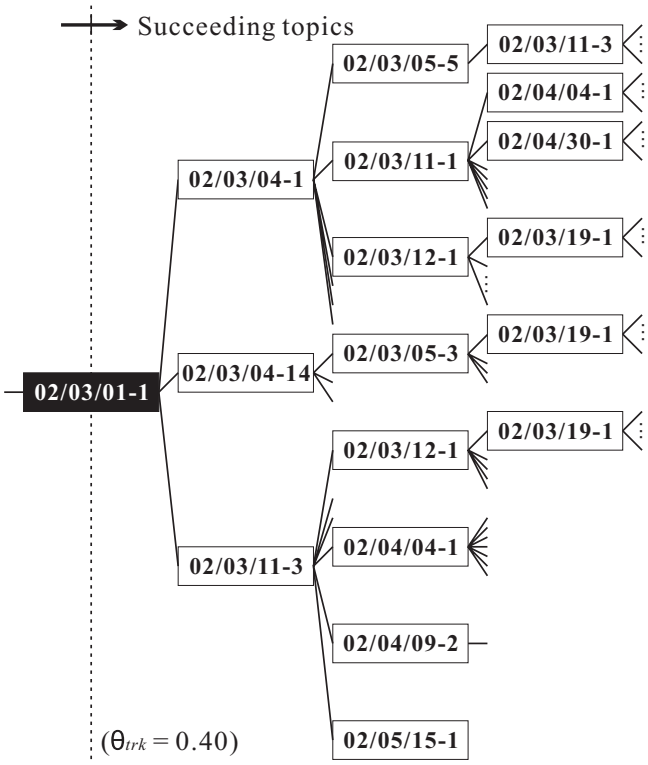

Figure 3: Example of topic structure without threading. Compare it with the right half of Figure 2.

2.2 Topic threading

When wandering through a very large-scale corpus, finding the right path to track down a topic of interest among numerous 'related' stories becomes a laborious work. In order to reduce the number of selections to the minimum, we propose a threading method that reorganizes the topic structure by pushing down stories that will eventually appear in the tracking process if necessary, or will never be selected anyway. Figure 3 shows the topic structure without threading. It shows that a non-threaded structure yields dozens of branches (candidates for selection) per tracking iteration, while the threaded structure in Figure 2 yields a relatively small number.

The thread structure is based on the combination of two aspects of relations between individual stories: 1) semantic relation and 2) chronological relation. Sequence of related stories are chronologically chained considering the two aspects to form a topic thread. Based on the nature of news stories, contents within a topic gradually transit along time, and subsequently a topic thread tends to merge and diverge with others. As a result, the corpus will be structured by complexly interwoven topic threads.

Topic tracking is realized by the following two steps.

1. Topic segmentation
2. Topic tracking/threading

Topic segmentation is performed within each video (*i.e.* intra-video structuring) referring to inter-sentence keyword vector relations. Topic tracking and threading is performed referring to inter-topic keyword vector relations throughout the entire video corpus. In the following Sections, the two steps are introduced in detail, followed by the visual interface based on the thread structure.

2.3 Related works

Topic segmentation and tracking is in general a part of the "Topic Detection and Tracking (TDT) task" defined by NIST [6], where various approaches have been proposed and evaluated in the past workshops and by related research community. On the other hand, there are several works that deal with news video corpora of a comparable size to ours such as [5] and the Informedia News-on-Demand project [8].

The approach proposed in this paper is significant especially in the tracking/threading scheme compared to these works, in that it pursues revealing highly structured relations and gradual transition of the topics.

3. INTRA-VIDEO STRUCTURING: TOPIC SEGMENTATION

3.1 Procedure

Topic boundaries are detected by applying the following procedure to daily closed-caption texts.

1. Apply morphological analysis to each sentence of a closed-caption text to extract compound nouns. A Japanese morphological analysis software, JUMAN [4] was employed. Compound nouns were extracted since combination of adjacent nouns was considered as more distinctive to represent a topic, rather than a group of individual nouns.

2. Apply semantic analysis to the compound nouns to generate a keyword frequency vector for each semantic class (general, personal, locational/organizational, or temporal) per sentence ($\vec{k_g}, \vec{k_p}, \vec{k_l}, \vec{k_t}$), which has frequencies as values. The analysis is done by a suffix-based method [2], which classifies compound nouns both with and without proper nouns, according to suffix dictionaries for each semantic class.

3. Set a window size w, and evaluate relations between w preceding and succeeding vectors at each sentence boundary. The relation at the boundary between sentences i and $i+1$ is defined as follows.

$$R_{S,w}(i) = \frac{\sum_{m=i-w+1}^{i} \vec{k_S}(m) \cdot \sum_{n=i+1}^{i+w} \vec{k_S}(n)}{\left|\sum_{m=i-w+1}^{i} \vec{k_S}(m)\right| \left|\sum_{n=i+1}^{i+w} \vec{k_S}(n)\right|}$$
$$(i = w, w+1, ..., i_{max} - w)$$

Here, $S = \{g, p, l, t\}$ and i_{max} stands for the number of sentences in a daily closed-caption text. We set $w = 1, 2, ..., 10$ in the following experiment.

4. Evaluate the following function.

$$R_S(i) = \max_w R_{S,w}(i)$$

The maximum of $R_{S,w}(i)$ along the w axis is taken. Preliminary observation showed that although most boundaries were correctly detected regardless of w, there was a large number of over-segmentation. Taking the maximum should mutually compensate for over-segmentations at various window sizes, since the following tendencies were observed in examples such as Figure 4.

- Small w (dotted line):
  Causes numerous over-segmentations, but has the advantage of showing significantly high relations within a short topic.
- Large w (dashed line):
  Does not always show high relations within a short topic, but shows relatively high relations within a long one.

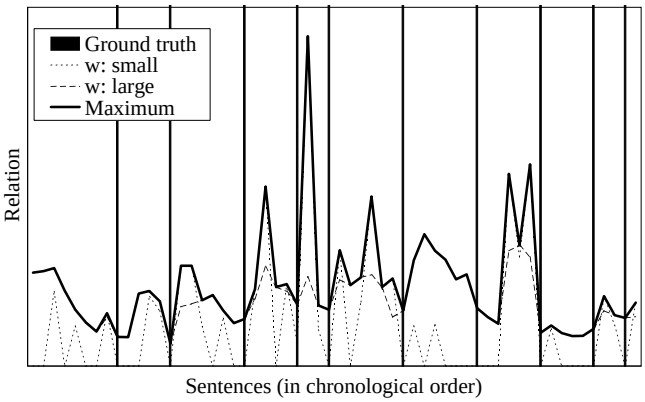

Figure 4: Over-segmentation tendencies according to window sizes.

The solid line, which represents the maximum of the relations evaluated in all window sizes shows the effect of this approach.

5. Evaluate the following function.

$$R(i) = \sum_{S=\{g,p,l,t\}} a_S R_S(i)$$

Relations evaluated in separate semantic attributes are combined as a weighted sum to evaluate the overall relation. Different weights are assigned to each semantic class under the assumption that especially in news texts, certain attributes should be more important than others when considering topic segmentation.

Multiple linear regression analysis was applied to manually segmented training data (consists of 39 daily closed-caption texts, with 384 boundaries), which resulted in obtaining the following weights.

$$(a_g, a_p, a_l, a_t) = (0.23, 0.21, 0.48, 0.08) \quad (1)$$

The obtained weights show that temporal nouns (*e.g.* today, last month) are not distinctive in the sense of representing a topic, where the other three, especially locational/organizational nouns act as distinctive keywords.

Finally, if $R(i)$ does not exceed a certain threshold θ_{seg}, the boundary between sentences i and $i+1$ is judged as a topic boundary.

Figure 5 illustrates the process through Steps 2. to 5.

6. Create a keyword vector $\vec{K_S}$ for each detected topic, and re-evaluate the relations between adjoining stories i and $j(= i+1)$ by the following function to concatenate over-segmented topics.

$$R(i,j) = \sum_{S=\{g,p,l,t\}} a_S \frac{\vec{K_S}(i) \cdot \vec{K_S}(j)}{\left|\vec{K_S}(i)\right| \left|\vec{K_S}(j)\right|} \quad (2)$$

As for a_S, the same weights as defined in Equation 1 were used.
If $R(i,j)$ does not exceed a certain threshold θ_{cat}, the adjoining topics are concatenated. This process continues until no more concatenation occurs.

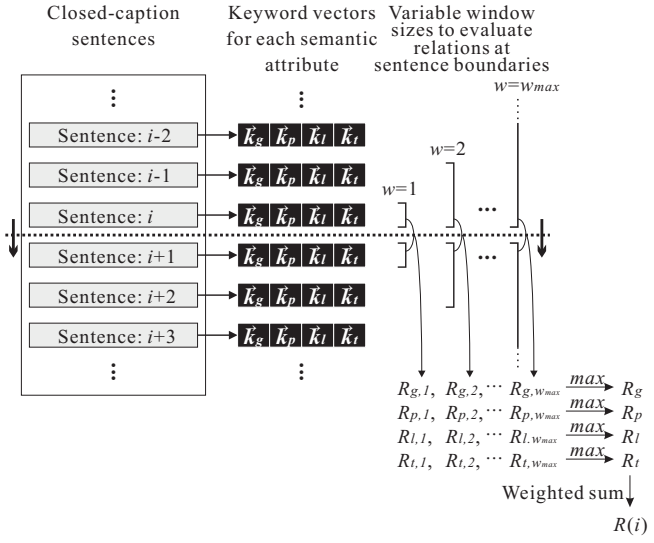

Figure 5: Evaluation of relations at sentence boundaries.

Table 1: Evaluation of topic extraction.

Condition	Both ends strict	One end strict/loose	Both ends loose
Recall	30.0%	34.6%	95.4%
Precision	28.5%	32.8%	90.5%

3.2 Experiment and evaluation

The procedure was applied first to the training data used in Step 4. to define appropriate thresholds ($\theta_{seg} = 0.28, \theta_{cat} = 0.08$), and later to the entire corpus ranging from March 16, 2001 to September 14, 2003 (826 days with 105,047 sentences in total) on a Sun Blade-1000 workstation with dual UltraSPARC-III 750MHz CPUs and 2GB of main memory. As a result, 12,005 topics with more than two sentences were extracted. Topics with only one sentence (26,740 topics) were excluded since they tend to be noisy fragments resulting from over-segmentation. The whole process takes approximately 5 seconds per day.

Evaluation was performed by applying the same procedure to manually segmented test data (consists of 14 daily closed-caption texts with 130 topics, set aside from all the training data used above), which resulted as shown in Table 1. Boundaries were counted as correct if they matched exactly with the manually determined ones in 'strict' condition, and allowing $\pm$ 1 sentences in 'loose' condition. The 'loose' condition is acceptable for our goal since sentences at true boundaries tend to be short and relatively less informative regarding the main news content.

4. INTER-VIDEO STRUCTURING: TOPIC TRACKING/THREADING

After topic segmentation, relations between all combinations of topics are evaluated for tracking by again adopting Equation 2. The weights a_S will be adjusted for interactive tracking in the future, though currently they are set to the same values as in Equation 1. When the relation $R(i, j)$ between stories i and j exceeds a certain threshold θ_{trk}, i and j are considered as highly related and linked for tracking. As the corpus grows, the amount of computation increases linearly to evaluate relations between daily increment and the original corpus. Currently this process takes up to 4 hours per day in order to compare an average of 14.5 new topics per day and more than 12,000 topics in the corpus.

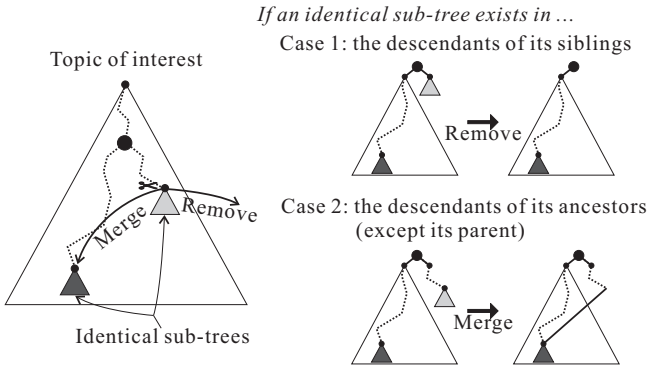

Figure 6: Topic threading scheme.

Following the evaluation of relations comes topic threading. The purpose of the threading is to form a chain of related stories along time in order to guide the users through the development of a topic of interest.

A topic thread structure originating from a topic of interest is formed by the following algorithm.

1. Create a tree originating from the topic of interest so that it satisfies the following conditions.
   (a) Children are directly tracked (highly related) stories linked to a parent, under the condition that their time stamps always chronologically succeed the parent's.
   (b) Siblings are sorted so that their time stamps always chronologically succeed their left-siblings'.
2. For each sub-tree $T_s(i)$, if an identical sub-tree $T_s(j)$ exists on the left-side, perform either of the following operations.
   (a) Remove $T_s(i)$, if $T_s(j)$ is a descendant of $T_s(i)$'s sibling, or else
   (b) Merge $T_s(i)$ with $T_s(j)$, if $T_s(j)$ is a descendant of $T_s(i)$'s ancestor (except for its parent).

   The sub-tree is removed in (a) instead of merging, to avoid creating shortcuts without specific meaning in the divergence and merger of the thread structure. The removal and merger scheme is shown in Figure 6. As a result of this operation, the thread structure will form a chronologically-ordered directed graph.

Note that this is in the case of forming a succeeding thread structure. Replace 'succeed' to 'precede' when forming a preceding thread structure.

In practice, the following conditions are applied to reduce computation time.

1. Perform Step 2. whenever an identical story is found during the expansion of the tree in Step 1. for pruning.

Figure 7: The "Topic Browser" interface.

2. Interrupt the expansion of the tree at a certain depth N_{trk}.

As the result of applying Condition 2., the resultant thread structure becomes an approximate one, though this will not affect much when referring to the direct children of the root (topic of interest) if N_{trk} is set to an appropriate value (We found $N_{trk} = 3 \sim 5$ as sufficient in most cases).

5. A TOPIC THREAD STRUCTURE-BASED USER INTERFACE

We implemented a topic browsing interface, namely the "Topic Browser" to provide the users with access to the topics based on the thread structure. It consists of two interfaces: the "Topic Finder" and the "Topic Tracker". Figure 7 shows the "Topic Browser" interface. The left side of the browser displays each interface (inter-switchable by selecting the corresponding tab). The right side displays the video and the closed-caption text corresponding to the topic selected in the tab.

The "Topic Finder" (Figure 8) is a portal to the topic browsing interface. First, a user inputs a query term. Then the interface returns topics that contain the query term in chronological order. Each topic is represented by a thumbnail image (the first frame of the video segment corresponding to the segment) and an excerpt of the closed-caption text. A user browses through them and selects the most relevant one to his/her interest to set it as an initial topic for the tracking process.

Next, the "Topic Tracker" (Figure 9) is an interface to track up and down a topic thread interactively. Although the initial topic should be selected through the "Topic Finder", the consecutive tracking can be done solely within this interface. The interface presents topic threads originating from the selected topic of interest in chronological order, separated in two categories: preceding topics and succeeding topics. Each thread is represented by the children of the selected topic of interest, along with thumbnail images of preceding/succeeding topics in the thread. Since the representative story could be considered as divergence points of topic threads, selecting a story of interest narrows down the

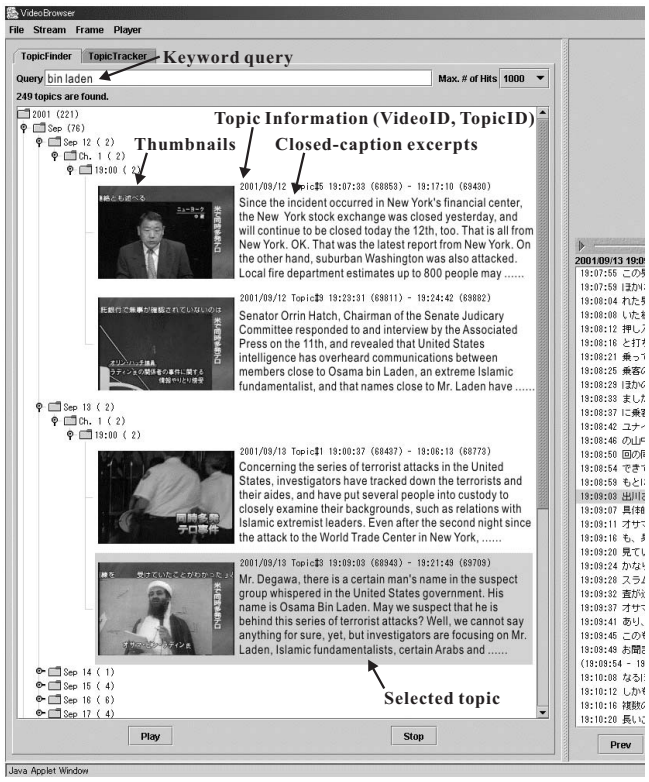

Figure 8: The "Topic Finder" interface (Texts originally in Japanese). Result of a query "Bin Laden".

subsequent tracking process.

The terms "preceding" and "succeeding" represent the chronological relations with the selected topic. A user could either track anterior or posterior sequence of topics by selecting one of the presented threads, and setting it as the next selected topic. Such interactive tracking goes on topic after topic, and the thread structure is re-analyzed on-the-fly after each interaction in order to reflect the user's intention. Figure 9 shows the list of topic threads starting with the topic selected in Figure 8. Each thread represents a different subject related to the series of terrorist attacks on September 11th and its aftermath.

The "Topic Finder" may seem somewhat similar to conventional keyword-based news video retrieval, but the combination with the "Topic Tracker" narrows down the result according to the users' interests and intentions. On the other hand, while narrowing down the results, the tracking could also be considered as a query expansion process following the gradual transition of the topic. Having the two seemingly opposite characteristics, the tracking process provides a user with a topic thread that matches their interests and intentions to the maximal extent. Moreover, it reveals chronological transition, divergence, and merger of topic threads, which will be effective for thorough understanding of the story related to the initial query. We found the above-mentioned features effective and informative after trying to track down several topics of interest. Figure 11 shows an example of a tracking tried out by one of the authors.

Similar approaches are taken in [1], although they fix the

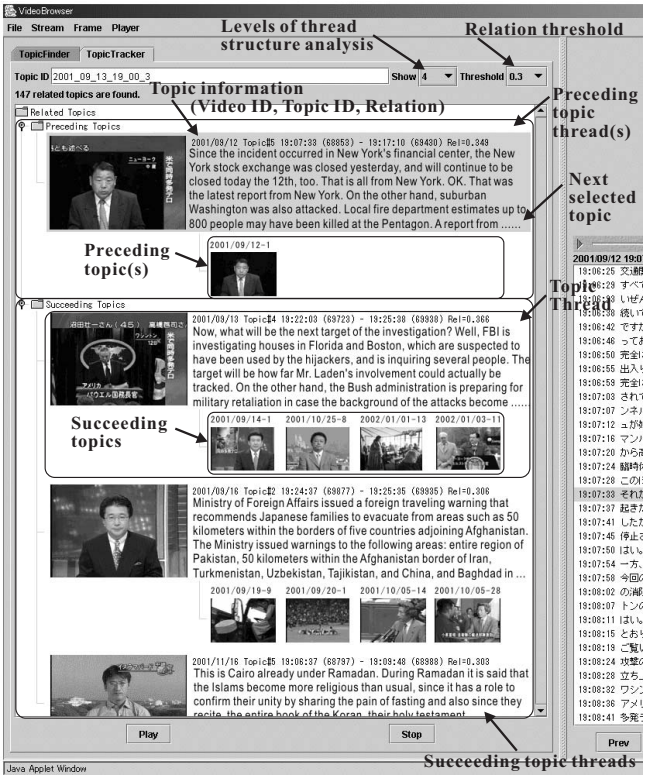

Figure 9: The "Topic Tracker" interface (Texts originally in Japanese). [Preceding topic(s)] Thread 1: Report on the series of terrorist attacks on September 11th. [Succeeding topics] Thread 1: Investigation on the attack; Thread 2: Travel warning by the Ministry of Foreign Affairs; Thread 3: Beginning of Ramadan in Egypt.

browsing interfaces within individual semantic attributes. On the contrary, our interface is designed to cope flexibly with various and vague interests and intentions of the users.

6. CONCLUSIONS

In this paper, we proposed a topic thread-based structuring method for a very large-scale news video corpus as a first step to reveal the underlying content-based structure. First, methods to segment and track/thread topics by closed-caption text analysis were described and evaluated. Next, interactive user interfaces that provide the users with the ability to browse through the corpus based on the topic structure were introduced. Although the proposed method is designed to handle Japanese news, it should be applicable to news in other languages if certain text processing components were to be provided.

We will investigate on achieving better segmentation and tracking/threading ability by integrating image-based structures to compensate for the limitation of text-based analysis such as lack of distinctive keywords. Topic segmentation will be improved by referring to image-based intra-video structure, for example, through anchor shot detection. Topic tracking will be improved by referring to graphically identical shots [7] based on the assumption that the same video footage is played over and over in news videos discussing the same topic, due to the limitation of data source or usage of referential material.

The amount for computation is a important issue to handle a large-scale corpus. since it will easily grow by squares in proportion to the size of the corpus. To cope with this issue, we will seek for approximate methods such as those introduced in the threading process. On the other hand, we also expect increase in computation power commercially available along with the expansion of the corpus.

Better tracking interface will be sought by providing thumbnails from more distinctive representative shots and representative keywords for each thread as introduced in [3]. The tracking process will also be improved by dynamically adjusting the weights used in Equation 2, depending on the user's initial query terms and tracking history in a relevance feedback manner. Such adjustment should provide the users with related topics reflecting their intentions.

7. ACKNOWLEDGMENTS

This work is partially supported by the Grants-in-Aid for Scientific Researches (15017285, 15700116) from the Japanese Society for the Promotion of Science, and the Japanese Ministry of Education, Culture, Sports, Science and Technology.

8. ADDITIONAL AUTHORS

Additional author: Shin'ichi SATOH (National Institute of Informatics, 2–1–2 Hitotsubashi, Chiyoda-ku, Tokyo 101–8430, Japan, email: `satoh@nii.ac.jp`).

9. REFERENCES

[1] M. G. Christel, A. G. Hauptmann, H. D. Wactler, and T. D. Ng. Collages as dynamic summaries for news video. In *Proc. 10th ACM Intl. Conf. on Multimedia*, pages 561–569, Dec. 2002.

[2] I. Ide, R. Hamada, S. Sakai, and H. Tanaka. Semantic analysis of television news captions referring to suffixes. In *Proc. 4th Intl. Workshop on Information Retrieval with Asian Languages*, pages 37–42, Nov. 1999.

[3] I. Ide, H. Mo, N. Katayama, and S. Satoh. Topic-based structuring of a very large-scale news video corpus. In *AAAI 2003 Spring Symposium on Intelligent Multimedia Knowledge Management*, Mar. 2003.

[4] Kyoto Univ. Japanese morphological analysis system JUMAN version 3.61, May 1999.

[5] A. Merlino, D. Morey, and M. Maybury. Broadcast news navigation using story segmentation. In *Proc. 5th ACM Intl. Conf. on Multimedia*, pages 381–391, Nov. 1997.

[6] National Institute of Standards and Technology. The 2002 topic detection and tracking (TDT2002) task definition and evaluation plan, May 2002.

[7] S. Satoh. News video analysis based on identical shot detection. In *Proc. 2002 IEEE Intl. Conf. on Multimedia and Expo*, volume 1, pages 69–72, Aug. 2002.

[8] H. D. Wactlar, M. G. Christel, Y. Gong, and A. G. Hauptmann. Lessons learned from building a Terabyte digital video library. *IEEE Computer*, 32(2):66–73, Feb. 1999.

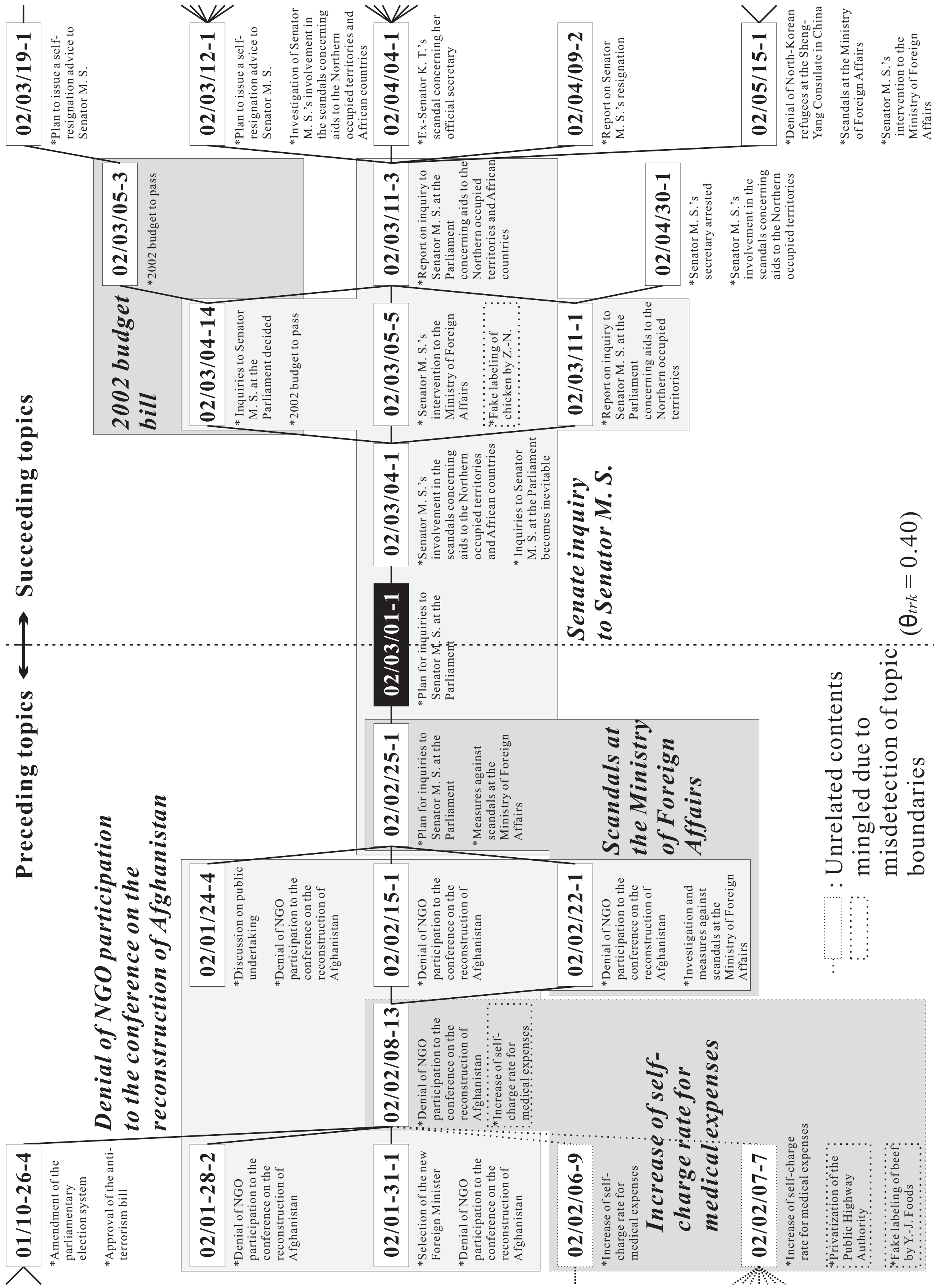

Figure 10: Details of topic threads extracted from the corpus. Topics are named in the following format: "Year/Month/Day-Topic#". Contents are annotated manually for reference.

September 12, 2001; Topic#1

In the United States, hijacked airliners slammed in the World Trade Center in New York, and the Pentagon in Washington on Tuesday. Rescue efforts are on the way at the sight, but the work is not proceeding smoothly. The death toll from the series of terrorist attacks could top several thousands. Ministry of Foreign Affairs has confirmed the safety of some 300 Japanese employees of 36 companies housed in the World Trade Center in New York, but there are still 18 unaccounted for. Last night, before 10 .....

September 12, 2001; Topic#5

Since the incident occurred in New York's financial center, the New York Stock Exchange was closed yesterday, and will continue to be closed today the 12th, too. That is all from New York. OK. That was the latest report from New York. On the other hand, suburban Washington was also attacked. Local fire department estimates up to 800 people may have been killed at the Pentagon. A report from Washington is by Kenji Sobata. Mr. Sobata? Yes. Is the Pentagon still burning? Yes, can you see the .....

September 13, 2001; Topic#3

Mr. Degawa, there is a certain man's name in the suspect group whispered in the United States government. His name is Osama Bin Laden. May we suspect that he is behind this series of terrorist attacks? Well, we cannot say anything for sure, yet, but investigators are focusing on Mr. Laden, Islamic fundamentalists, certain Arabs and Middle Easterns, and so on. Osama Bin Laden is a leader of Islamic fundamentalists, and he is said to be in the back of an international network. We interviewed an Egyptian professional .....

September 13, 2001; Topic#4

Now, what will be the next target of the investigation? Well, FBI is investigating houses in Florida and Boston, which are suspected to have been used by the hijackers, and is inquiring several people. The target will be how far Mr. Laden's involvement could actually be tracked. On the other hand, the Bush administration is preparing for military retaliation in case the background of the attacks become clear. Secretary of States, Collin Powell stated that diplomatic consensus is becoming formed among .....

September 14, 2001; Topic#1

The United States government says that at least 18 people were involved in the attacks, and it is becoming increasingly convinced that an Islamic Fundamentalist leader, Osama Bin Laden was behind the attacks. The Bush administration has said it is planning to launch comprehensive military retaliation for the attacks against the terrorist organizations responsible and any nation that supports them. I'm looking at those terrorist organizations, who have the kind of capacity that would have been .....

September 15, 2001; Topic#1

Good evening, it is 7 PM, Saturday September 15th. Tonight's program will be extended to 8 o'clock. We have extensive coverage of the terrorist attacks in the United States. The United States Congress has approved the resolution allowing the Bush administration to use force to retaliate against Tuesday's terrorist attacks. President Bush is preparing seriously for the military retaliation to the terrorist organizations. The resolution allows full-measure military attacks to terrorist organizations .....

Figure 11: Detailed example of topic tracking: The thread was selected by one of the authors (Texts are originally in Japanese). Starting from the report on the attack, the process of the investigation and the path to the attack to Afghanistan could be understood.