Content Augmentation and Webification for Enhancing TV Viewing

Qiang Ma[1], Hisashi Miyamori[1], and Katsumi Tanaka[1,2]

[1] National Institute of Information and Communications Technology,
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0289, Japan
{qiang, miya}@nict.go.jp
[2] Graduate School of Informatics, Kyoto University,
Yoshida Honmachi, Sakyo, Kyoto, 606-8501 Japan
tanaka@dl.kuis.kyoto-u.ac.jp

Abstract. A system is described for enhancing the viewing of programs on storage televisions. The content of a program is webified and augmented by analyzing the closed captions to structuralize the content online and by searching for Web pages that provide information complementary to the program. The structuralized content and related information are viewed using an intuitive, zooming user interface that enables the user to switch gradually from watching a program to browsing the program like a Web page and to change the level of detail. Prototype testing validated the concept of this "**WA-TV**" (**W**ebifying and **A**ugmenting **TV**-content) system.

1 Introduction

Constant advances in information technologies and the spread of these technologies have altered our daily lives considerably. For instance, digital broadcasting and storage television combining broadcasting and computer technologies are changing the way we watch television. While television programs can provide excellent quality and realism, they suffer from restrictions on time and an obligation to accommodate popular opinion, which limit the level of detail and the scope of information they can provide. In contrast, information published via the Internet is diverse and faces few restrictions. Thus, there is a great need for functions that can provide additional information about the TV programs in which we are interested.

In addition, the introduction of storage TV enables more than 1000 hours (about 600 GB) of programming to be recorded at a certain level of quality. This is changing the recording style from "searching-recording" to "recording-searching". In other words, instead of searching a program guide for programs to record, we can now record a great many programs and then later search for interesting ones to watch. However, since we do not have an unlimited amount of time to spend searching through a great amount of content, there is a great need for functions that enable particular video segments to be selected from a huge amount of recorded data, that present an overview of the segment content in a compact form, and that can provide a digest of the content in a limited amount of time.

E.A. Fox et al. (Eds.): ICADL 2005, LNCS 3815, pp. 129–138, 2005.

We previously proposed a primitive version of an application system we call ”WA-TV” (Webifying and Augmenting TV content) that works offline for browsing video content[6]. Actually, in Japanese, ”WA” can mean ”Japan”, ”fusion”, ”harmonious”, and ”smooth”. We have now extended this system to work online and supplement viewing of storage television. We use online text stream segmentation and complementary information retrieval methods. In contrast to conventional TV viewing, our system provides additional information about a program being watched by online analyzing the text stream (closed captions, etc.) and retrieving information from the Internet. It enables gbrowsingh a TV program like a Web page and switching gradually from TV watching to TV browsing, enabling the user to explore video segments of interest.

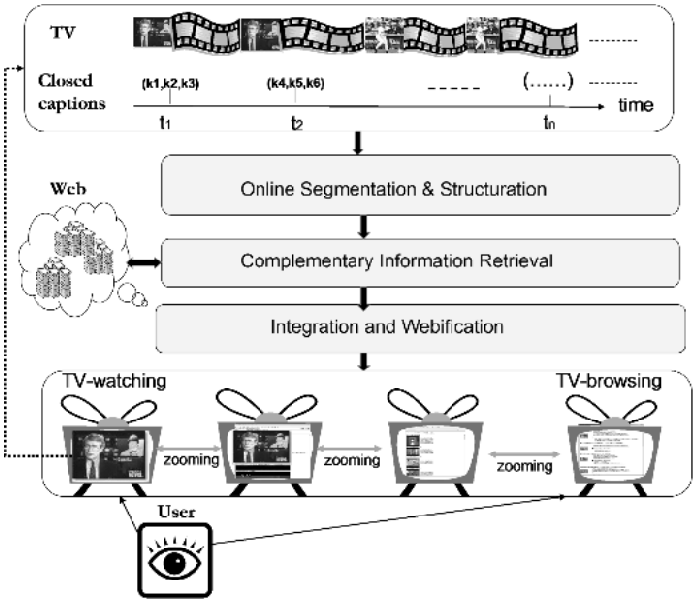

Fig. 1. Overview of WA-TV

In this paper, we assume that closed captions are broadcast continuously during a TV program via datacast. As illustrated in **Fig. 1**, WA-TV first analyzes the closed captions of a program and uses them to segment the scenes and to construct a hierarchical structure of the programfs topics. It then searches for Web pages complementary to each topic by using a complementary information retrieval method described elsewhere[4]. The segmented closed captions and corresponding scenes are grouped into pairs and laid out in the form of a storyboard. The retrieved complementary information is integrated at the corresponding positions in the storyboard. The display of this integrated information is controlled using a zooming interface. The sizes of the displayed images of the segmented scenes can be changed smoothly, and the storyboard can be switched to another one with a different level of detail. Users can thus seamlessly move back and forth between storyboard screens (TV browsing) with different levels of detail and

the normal playback screen (TV watching), enabling them to easily explore for specific scenes. Moreover, hyperlinks to the related information are integrated into each storyboard, so users can efficiently access the related information.

In Sect. 2 we discuss related work. In Sect. 3 we describe the online segmentation methods used for structuring TV programs. The complementary information retrieval method is described in Sect. 4. In Sect. 5 we describe the zooming interface. In Sect. 6 we show some experiment results. We conclude with a brief summary and our plans for future research in Sect. 7.

2 Related Work

A lot of research has addressed the display of video overviews and the creation of video digests[1, 9, 11]. These methods improve browsability or comprehension of content in a limited time by spatially or temporally expanding key frames or video segments. They basically reduce the amount of information displayed to the user. In contrast, WA-TV augments the information displayed to users via hyperlinks, while at the same time improving browsability and content comprehension in a limited time.

Informedia[2, 3] introduced various methods for video segmentation based on analysis of closed captions. However, because these methods require scanning of the whole body of data, they cannot be applied to data streams, which are received continuously.

Henzinger et al. proposed methods for automatically generating queries from closed captions that can be used to find Web pages with content similar to that of the program being watched[7]. Unlike their approach, our mechanism does not search for Web pages with content merely similar to that of the program. It searches for pages that provide additional information.

3 Online Structuration of TV Programs

In contrast to conventional text stream segmentation methods[2, 3, 10], of which most are top-down approaches, we propose a bottom-up segmentation method that incrementally identifies the story boundary so that it does not need to scan all the data. The basic unit used for further processing is the closed captions received at a certain time. We call such unit block. In the closed captions for a Japanese news program, for example, one sentence generally straddles more than two blocks. Our method for online

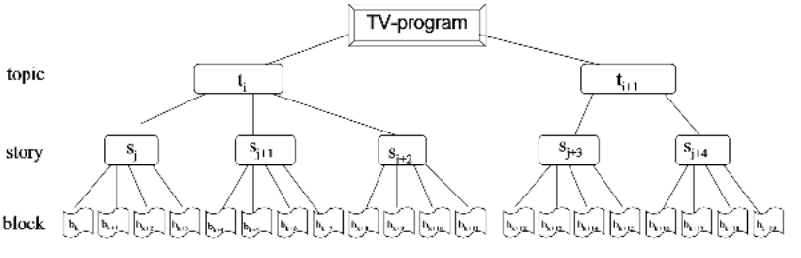

Fig. 2. Hierarchical Structure of TV Program

structuring of TV programs includes two phases: 1) story segmentation and 2) merging related adjacent stories into one topic. The result is a hierarchical structure of a program corresponding to its closed captions: topic, story, and block, as shown in **Fig. 2**.

3.1 Incremental Story Segmentation

The basic idea is that a high rate of keyword pairs with strong co-occurrence relationships among all keyword pairs within various closed captions suggests that these captions describe one story. Intuitively, we compute the co-occurrence relationships of keywords in the received closed captions data. If their co-occurrence relationships are strong, the corresponding closed captions may describe the same story. We then merge them with the next set of closed captions and recompute the keyword co-occurrence relationships. If they are weak, 'noisy' captions describing another story may have been received, so there should be a boundary identifying the story change.

When words w_1 and w_2 co-occur frequently within a text collection, we say that they have a strong co-occurrence relationship and that their co-occurrence rate is high. In this paper, we estimate the co-occurrence rate $cooc(w_i, w_j)$ between the words w_i and w_j as follows.

$$cooc(w_i, w_j) = \frac{df(\{w_i, w_j\})}{df(\{w_i\}) + df(\{w_j\}) - df(\{w_i, w_j\})} \tag{1}$$

where $df(\{w_i\})$ is the number of texts containing the word w_i within a pre-specified text collection, and $df(\{w_i, w_j\})$ is the number of texts containing both w_i and w_j.

The details of the procedure are as follows (see also **Fig. 3**). Here, CT_i is the keyword set used to detect a story at time point t_i. ST and ET, respectively, are the initial and terminal time points of an identified story.

1. Let $CT_0 = \emptyset, ST = 0, i = 1$.
2. Receive closed captions. If there are no other closed captions, stop.
3. After receiving the closed captions at time point t_i ($i \geq 1$), extract keyword set K.

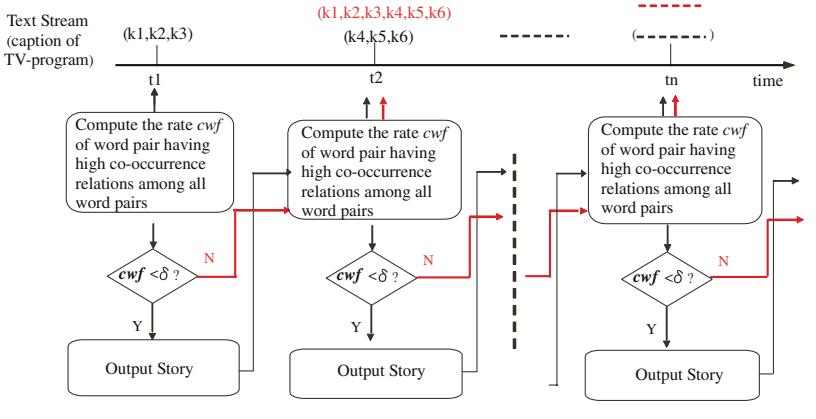

Fig. 3. Online Story Segmentation (t_i is the time point at which closed captions are received.)

4. Let $CT_i = CT_{i-1} \cup K$.
5. Compute rate $cwf(t_i)$ of keyword pairs with high co-occurrence rates ($\geq \theta$) among all keyword pairs within CT_i. Here, θ is a pre-specified threshold, and m is the number of keywords included in CT_i.

$$cwf(t_i) = \sum_{j=1,k=j+1}^{j=m-1,k=m} cr(w_j, w_k) / \frac{m \cdot (m-1)}{2} \tag{2}$$

$$cr(w_j, w_k) = \begin{cases} 1 \text{ if } cooc(w_j, w_k) \geq \theta \\ 0 \text{ if } cooc(w_j, w_k) < \theta \end{cases} \tag{3}$$

6. If $cwf(t_i) > \Theta$, go to 9. Θ is a pre-specified threshold.
7. Let $ET = t_i$. The initial and terminal time points of output $story_i$ are ST and ET, respectively. The keywords for the story are also output for further processing.
8. Let $CT_i = \emptyset, ST = t_i$.
9. $i = i + 1$. Receive closed captions. If there are no other closed captions and $CT_i = \emptyset$, stop. If there are no other closed captions, and $CT_i \neq \emptyset$, go to 7. Otherwise, go to 3.

3.2 Topic Segmentation by Incremental Joining of Stories

We try to merge the story identified using the above method with the next story and continue doing so until a merger can no longer be done. The merger of stories is based on join of their topic structures[**?**, 5]. Here, we give a brief overview of the topic structure model and its joining operation.

Intuitively, a topic structure consists of a pair of subject and content terms. The subject terms denote the dominant terms on a Web page or in a text stream (keyword sequence, e.g., closed captions for videos). A content term is a term that has strong co-occurrence relationships with the subject terms. In other words, subject terms are centric keywords that play a title role on a Web page (or video), and the content terms play a supporting (or describing) role. The subject and content terms are extracted by using the term frequency (tf) and the co-occurrence relationship between two terms. In short, if a keyword has high rates of co-occurrence with other keywords and its term frequency is higher than that of other keywords, it is considered to be a subject term. Of the remaining keywords, those that have a high co-occurrence relationship with the subject terms have a high probability of being content terms.

A topic structure is represented as a connected directed acyclic graph (DAG) called a topic graph. In a topic graph, a node denotes a keyword, subject term, or content term. A directed edge denotes the subject-content relationship between two keywords. The join of two topic structures, s and s', is defined as the union of their topic graphs.

$$s \bowtie s' = \begin{cases} G(s) \cup G(s'), \text{ if } G(s) \cup G(s') \text{ is a connected DAG.} \\ \phi, \text{ otherwise} \end{cases} \tag{4}$$

where $G(s)$ and $G(s')$ stand for the respective topic graphs of s and s', and ϕ stands for null. In addition, $s \bowtie \phi = \phi$.

If the result of joining the topic structures of two stories is not ϕ, they are merged. All stories that can be merged together are organized into one topic. In other words, a topic is a series of related stories that can be merged based on join of their topic structures.

$$\begin{aligned} topic = s_i \bowtie s_{i+1} \bowtie ... \bowtie s_j \\ s_i \bowtie s_{i+1} \bowtie ... \bowtie s_j \neq \phi \\ s_i \bowtie s_{i+1} \bowtie ... \bowtie s_j \bowtie s_{j+1} = \phi \end{aligned} \tag{5}$$

where s_i is the initial story of this topic. Obviously, s_{j+1} is the initial story of the next topic, in the given definition.

4 Structured Queries for Complementary Information Retrieval

We defined four types of queries for finding Web pages related to the given story and topic: 1) CD (content-deepening), 2) SD (subject-deepening), 3) SB (subject-broadening), and 4) CB (content-broadening) queries[4].

CD and SD queries are based on a join such that the subject terms in one topic structure appear as the content terms in the other. Such joins add more detail to the original information. SB and CB queries are based on a join such that two topic structures have the same subject or content terms. Such joins provide broader coverage of the information.

A previous report [8] showed the feasibility of extracting the topic structures of a Web page by using the "title" and "body" tags. Based on this work, we assume that the keywords appearing in the title and body of a Web page are its subject and content terms, respectively. Thus, we can use the structure option of Google, intitle, intext, etc., to search for Web pages complementary to the TV program.

Hereafter, let topic structure t of the segmented story be $(\{s_1, s_2\}, \{c_1, c_2, c_3\})$, where s_i and c_i stand for a subject term and content term, respectively. "intitle" and "intext" indicate that the following terms are the respective subject and content terms of a topic structure contained in the retrieved Web page. "$\wedge$" and "$\vee$" stand for "logical AND" and "logical OR", respectively. "$\neg$" means "logical NOT".

1. Content-Deepening Queries:

$$(intitle : c_1 \wedge c_2 \wedge c_3) \wedge (\neg(intext : s_1 \vee s_2)) \tag{6}$$

2. Subject-Deepening Queries:

$$(intext : s_1 \wedge s_2) \wedge (\neg(intitle : c_1 \vee c_2 \vee c_3)) \tag{7}$$

3. Subject-Broadening Queries:

$$(intext : c_1 \wedge c_2 \wedge c_3) \wedge (\neg(intitle : s_1 \wedge s_2)) \tag{8}$$

4. Content-Broadening Queries:

$$(intitle : s_1 \wedge s_2) \wedge (\neg(intext : c_1 \wedge c_2 \wedge c_3)) \tag{9}$$

We issue these queries to Google. The top result of each query is regarded as the complementary Web page and will be integrated with the corresponding story (or topic).

5 Zooming User Interface

The structured program data (topic, story, and block) is integrated with the retrieved related information into Web content for display. An example snapshot displayed on WA-TV is shown in **Fig. 4**. The segmented caption texts and videos are displayed vertically in the form of a storyboard. Hyperlinks to the complementary information are located below the caption texts, enabling users to access more detailed or broader information than provided by the broadcast program. The transformation of the screen appearance is illustrated in **Fig. 5**. The zooming feature can be used to smoothly change the size of the thumbnails as well as to switch from one storyboard to another with a different level of detail.

For example, suppose that we are watching on TV. Zooming-out smoothly switches the watching (playback) scene to a storyboard including one block. Further zooming-out smoothly change the size of TV-viewer on the storyboard, and when the size reaches a certain level, the storyboard switches to one containing thumbnails of stories (current story and previous stories) of TV-program. The corresponding closed captions and hyperlinks to complementary Web pages are also displayed. Further zooming out will smoothly changes the size of the thumbnails on the storyboard, and when their size reaches another certain level, the storyboard switches to one including thumbnails of topics (current topic and its previous topics) and related information (closed captions and links to complementary Web pages). Zooming-in produces the opposite effect.

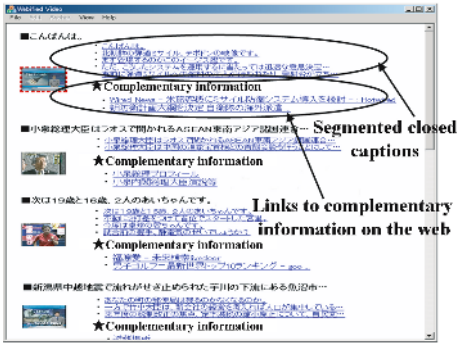

Fig. 4. Example Snapshot Displayed on WA-TV

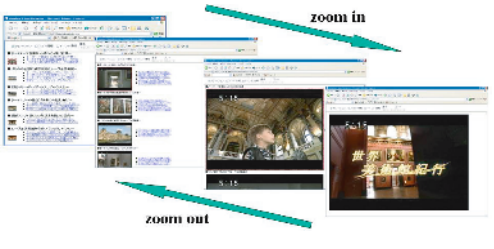

Fig. 5. Gradual Changing of TV Viewing On WA-TV

During the zooming operation, it is also possible to change the focus onto specific scene (story or topic). As a result, users can seamlessly move back and forth between storyboard screens showing different levels of detail and the normal playback (watching) screen, enabling them to easily explore for specific scenes.

6 Evaluation

6.1 Online Structuration Experiment

We used closed captions (in Japanese) from NHK News 7 (a well-regarded news program in Japan) collected over a 28-month period to build a co-occurrence relationship dictionary. We used ChaSen (chasen.aist-nara.ac.jp/) for Japanese morphology analysis and only nouns as keywords for further processing. To exclude stop words, we built a stop-word dictionary containing 593 terms in English and 347 terms in Japanese.

A boundary is defined as correct if and only if it is a true boundary. However, due to our use of topic-structure-based complementary information retrieval in WA-TV, one segmentation method usually produces satisfactory results because it always comes close to the true boundary. Here, we relax our correctness criteria to accept all boundaries that are one block off the true boundary. The distance between the identified boundary and the closest true boundary is defined as the degree of relaxation. **Fig. 6** illustrates the relaxed failure model for our block-based segmentation method adapted from Hauptmann et al. [3].

We used the closed captions for NHK News 7 programs collected over two days (821 blocks) as the experimental data. **Table 1** shows the experimental results. The reference boundaries for topics and stories were specified by evaluators beforehand. The *F-measure values* indicate that the proposed structuration method performed better

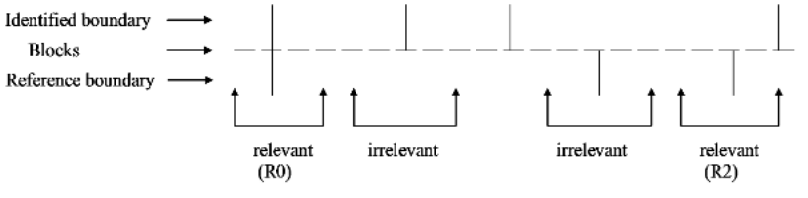

Fig. 6. Relaxed Failure Model of Block-based Text Segmentation Method (Rx means under the degree of relaxation x, identified boundary is OK.)

Table 1. Results of Online Structuration Experiment (Degree of Relaxation is 1)

	Story Segmentation			Topic Segmentation		
	Recall	Precision	F-measure	Recall	Precision	F-measure
$\theta = 0.2, \Theta = 0.2$	0.286	0.395	0.332	0.301	0.446	0.360
$\theta = 0.2, \Theta = 0.25$	0.426	0.330	0.372	0.360	0.387	0.373
$\theta = 0.25, \Theta = 0.2$	0.425	0.287	0.343	0.342	0.333	0.338

than the topic change detection method used in Informedia whose best F-measure value is 0.367[3].

6.2 Structured Query for Complementary Information Retrieval Experiment

We used 88 topic structures extracted from the closed captions collected over three days for NHK News 7 programs. Each one consisted of two subject terms and three content terms.

For the CD, SD, CB, and SB queries, we used the top-ranked result from a Google search as the complementary Web page. Based on human assessment of the relevance of these complementary Web pages, the calculated precision ratios were 0.489, 0.625, 0.511, and 0.705, respectively. Details of the experimental results are shown in **Table 2**. When the query was based on a topic structure containing proper nouns, the search results were better. This suggests that proper nouns play an important role in using topic structures to retrieve information from the Web. We will examine this feature further in future work.

Table 2. Results of Complementary Information Retrieval Experiment

	topic structure	relevant pages	precision ratio
SB	88	62	0.705
SD	88	55	0.625
CB	88	45	0.511
CD	88	43	0.489

6.3 User Interface Evaluation

Simple experiments were conducted to evaluate the user interface, especially the zoom operation. Most of the participants (8 out of 11) found the ability to search for scenes by looking through a list of closed captions and thumbnails on the storyboard ”useful” compared to a conventional interface based on fast-forwarding and rewinding. They also found the ability to control the different levels of detail ”intuitive”. Evaluation tests using more participants will be conducted to better evaluate usability, understandability, etc.

An important advantage of WA-TV is that it enables active browsing of TV programs, which are conventionally viewed by passive watching, by converting them into Web content. WA-TV enables active browsing by hyperlinking various positions in the program to external related information with greater detail or from different perspectives.

7 Conclusion

We have described an application system for augmenting and webifying TV content to enhance viewing of storage televisions. WA-TV segments and structures a TV program into different levels of detail online and then generates hyperlinks between various positions in the program and complementary Web pages it retrieves to provide additional

information about the program. The retrieved information and the original TV content are integrated into a Web form for browsing. In addition, a zooming feature enables the user to switch gradually from TV watching to TV browsing. Experimental results valided the proposed methods.

We plan to improve the proposed methods used in WA-TV, particularly the online topic segmentation and complementary information retrieval. Further experiments with a larger number of participants are also planned.

References

[1] Christel, M.G. and Huang, C. Enhanced access to digital video through visually rich interfaces, *Proc. of ICME 2003*, 2003.

[2] H. D. Wactlar. Informedia - search and summarization in the video medium. In *Proceedings of Imagina 2000 Conference*, 2000.

[3] Huauptmann A., Chang J.C., Hu N.N., and Wang Z.R. Text Segmentation in the Informedia Project, *http://www-2.cs.cmu.edu/ hnn/project/ML-project/ml-report.htm*.

[4] Ma Q. and Tanaka K. Topic-structure-based Complementary Information Retrieval and Its Application, *ACM Transactions on Asian Language Information Processing (to appear)*, 2005.

[5] Ma Q. and Tanaka K. Topic-structure-based complementary information retrieval for information augmentation, *Proc. of APWeb2004, LNCS3007*, pp. 608-619, 2004.

[6] Miyamori, H., Ma, Q., and Tanaka K. WA-TV: Webifying and Augmenting Broadcast Content for Next-generation Storage TV, *Proc. of ICME2005*, 2005.

[7] Henzinger M., Chang B.-W., Milch B., and Brin S. Query-free news search. *Proc. of WWW2003*, 2003.

[8] Oyama S. and Tanaka K. Exploiting document structures for comparing and exploring topics on the Web. *Proc. of WWW2003 (poster tracks)*, 2003.

[9] Sumiya, K., Munisamy, M., and Tanaka, K. TV2Web: generating and browsing Web with multiple LOD from video streams and their metadata, *Proc. of ICKS2004*, pp. 158- 167, 2004.

[10] TDT site. http://www.itl.nist.gov/iaui/894.01/tests/tdt/index.htm, 2005.

[11] Uchihashi, S., Foote, J., Girgensohn, A., and Boreczky, J. Video Manga: generating semantically meaningful video summaries, *Proc. of ACM Multimedia 99*, 1999.