Complementing Your TV-Viewing by Web Content Automatically-Transformed into TV-program-type Content

Akiyo Nadamoto
National Institute of Information and Communications Technology
Hikaridai, Seikachyo, Kyoto, Japan
nadamoto@nict.go.jp

Katsumi Tanaka
Kyoto University
National Institute of Information and Communications Technology
Yoshida Honmachi, Sakyo-ku, Kyoto, Japan
ktanaka@i.kyoto-u.ac.jp

ABSTRACT

Despite much talk about the fusion of broadcasting and the Internet, no technology has been established for fusing web and TV program content. In this paper, we propose ways to transform web content into TV-program-type content as a first step towards the fusion of these media. Our transformation method is based on two criteria - the transmitted information and the dialogue among character agents. The method deals with both an audio component and a visual component. By combining these techniques, we can transform web content into various forms of TV-program-type content depending on the user's aims. We present three different prototype systems, u-Pav which reads out the entire text of web content and presents image animation, Web2TV which reads out the entire text of web content and presents character agent animation, and Web2Talkshow which presents keyword-based dialogue and character agent animation. These prototype systems enable users to watch web content in the same way, they watch a TV program.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.5.2 [**User Interface**]: Prototyping; I.7.m [**Document and Text Processing**]: Miscellaneous

General Terms

Design, Documentation

Keywords

Media fusion, Media conversion, web content, TV program content

1. INTRODUCTION

The Internet use is now very widespread and content such as movies can be easily downloaded from the Internet at home. The increasing popularity of the Internet has coincided with a major change in the broadcasting environment as digital broadcasting has been introduced and data broadcasting services have become available. With these advances, the fusion of broadcasting and the Internet has become a widely discussed topic. In the near future, these two media will have similar features and the boundaries between them will become increasingly blurred. How will this affect their content? It will become possible, for example, for consumers to create their own TV program content, which at present is created only by broadcasting companies. This means that content authors, who will include both professionals and consumers, will be able to send information using their favorite media, content style, and environment. Users will also be able to get information provided in their favorite content style. For example, they may want to watch web content in the same way they watch TV or browse TV programs as if they were browsing the web. They might also want to obtain TV program content and web content simultaneously; that is, the two forms of content will be blended and their boundaries between them will disappear. We regard the fusion of TV and web content as a key factor that will shape next-generation systems for content delivery. Both broadcasting and the Internet are reaching maturity, and the two environments both offer a wide range of content. However, one problem in developing methods too enable the fusion of existing content is that TV program content and web content typically have different structures. TV program content is time-based content and consists of movies (animation) and sound. In contrast, web content is two-dimensional, windows-based content that consists almost entirely of text and images. How can these different content structures be combined automatically? Here, we consider three ways of presenting fused TV program and web content(show Figure 1). In this paper, we refer to the new content created by fusing TV program and web content as "fusion content" and that created by transforming web content into TV-program-type content as "transformed content".

- TV-style presentation
  When users want to watch fusion content like they watch TV, a system automatically transforms web content into TV-program-type content. The system combines transformed content and TV program content, and then automatically creates fusion content that looks like TV program content.

- Web-browser-style presentation

MM'05, November 6–11, 2005, Singapore

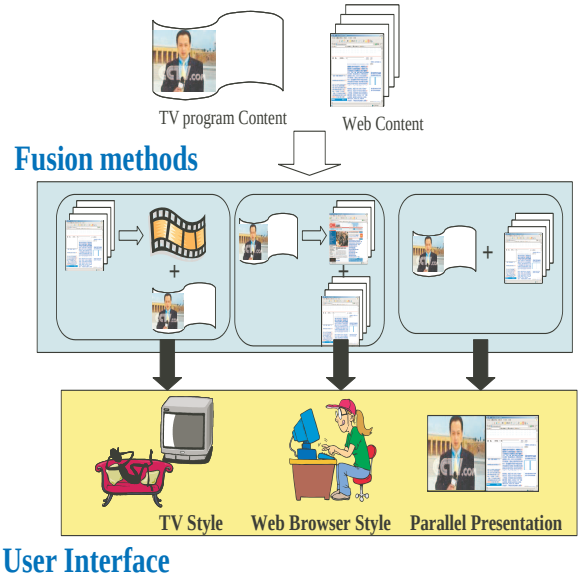

Figure 1: Three ways to present fusion content

When users want to browse fusion content like they browse the web, a system first segments TV program content into several passages and transforms the TV program content into web content based on movie data and meta data. The system then inserts the web content in the transformed content.

- Parallel Presentation
  With this method, TV program content and web content are presented simultaneously, enabling users to browse web content while watching TV program content.

Another of our projects involves research on presenting content in web-browser-style and parallel presentation, e.g. Miyamori et al.[6] have studied the use of a web-browser style for presenting content, and Ma et al.[13] have studied parallel presentation. Our focus is on methods for presenting content in a TV style.

We have described a simple automatic transformation technique [9]. In this paper, we describe the automatic transformation of web content into TV-program-type content from the viewpoint of the fusion of web and TV program content, and describe two criteria for this fusion - the transmitted information and the dialogue among character agents. Furthermore, we propose a new transformation technique which is based on automatically transforming web content into dialogue-based TV-program-type content.

The remainder of this paper is organized as follows: Section 2 discusses related work, Section 3 explains the basic concept of transforming web content into TV-program-type content, Section 4 explains how dialogue is created, Section 5 describes our prototype systems, and Section 6 shows the results from user evaluations. We conclude in Section 7.

2. RELATED WORK

Push-type information dissemination

ANATAGONOMY [14] is a push-type information dissemination system, which provides time-based information like TV broadcasts. Users can subscribe to the channels (subchannels) they are interested in, much like personalized newspapers. The interfaces for browsing the delivered information are, however, character-based, and users are forced to read each article one by one.

There are various kinds of RSS readers, e.g. Sharp Reader[2] and Headline Reader[1], which browse text-based news automatically, but they differ from our approach in that they do not transform web content into TV-program-type content.

TV and the Web

WebStage[15] provides a way of transforming web content into news-program-like content based on a TV-program metaphor described by FRIEND21[7]. Though it is similar to our system, our work is based on separating audio component and visual component, and we create various TV-program-type content based on the type of audio and visual component. WebTV[5] enables users to access the Internet through a TV without a computers. Users can send/receive e-mail and browse the Internet while also watching TV. In addition, web pages related to the on-air TV program are displayed automatically. WebTV combines several TV display and input-device techniques such as gray-scale fonts, alpha-blending, and a simple remote-control interface. This system provides a possible method for fusion of TV broadcasting and computers. Web content, however, still has to be browsed using conventional reading and clicking operations. The W3C has a working group which has described authoring techniques that enable device independence[4]. Their approach is to use authoring techniques based on a markup language, such as XML-based languages. In contrast, our approach is to automatically transform web content into TV-program-type content. This is the main point of difference. Our approach focus on dividing audio and visual components is very similar to their approach. MyInfo[8] processes and combines personal news applications from TV and the web. It also combines TV and web content in the same window. This research is aimed at finding ways of combining TV and web content, though, while our research is aimed at the transforming of web content into TV-program-type content.

TVML

TVML (TV program Making Language)[3], proposed by NHK Science and Technical Research Laboratories, is a tool for producing an entire TV program on a desktop. It is a kind of scripting language that can be used to describe computer graphics(CG)-based TV programs. A TV program script written in TVML is played like a conventional TV program by a TVML player; that is, a TVML script is translated into CG animation with synthesized speech, virtual camera movement, and real video. TVML was originally aimed at providing a framework for producing TV programs and was not intended for web browsing. However, since our goal is to transform web content into TV-program-type content and to fuse web content and TV program content so that it can be presented in a TV-style. TVML is a language that we might be able to use to develop systems for transforming web content into TV-program-type content.

Dialogue

There are many researches about Dialogue analysis. Ishizaki et al.[11] has provided a good summary of the work done in this area. In most cases, the approach has been to analyze real world dialogues and extract the intention from the

dialogue. Our research, aimed at transforming web content into dialogue-based TV-program-type content, takes the opposite approach.

3. BASIC CONCEPT

3.1 TV-style presentation of fusion content

When we watch TV, we sometimes have difficulty understanding what is said, or we would like more information about it. In these situations, it would be convenient to have a system that automatically presents related web content, as in the following scenario:

1. A user is watching a news program in real time while also recording it, or is watching a recorded news program on a TV.
2. The system is carrying out a background search for related web pages in real time.
3. The system stops presenting the news program.
4. The system presents the related web pages it has found.
5. After presenting the web pages, the system represents a sequel to the news program.

In this scenario, two technical problems have to be dealt with: (1) how to extract related web pages, and (2) how to present them. Ma et al.[12] have investigated the first issue. Therefore, the focus of this paper is on how to present web pages while the user is watching TV. Naturally, it is not easy to watch a TV program and browse web content simultaneously. We believe that presenting TV-program-type fusion content, that is, combining a TV program with related web content in the form of audio-visual content, is more practical.

3.2 Criteria of transformed content

Figure 2 shows tow types of criteria of our transformed content.

- **The transmitted information**
  Web content consists of text and images within the layout of the web page. We refer to the text and images from web content as "assets" and information regarding the page layout as "style". TV program content, on the other hand, consists of audio and visual information. We transform web content into separate audio and visual components. The web content style involves the layout of text and images for a two-dimensional window (i.e., a browser window). We do not apply this style to TV-program-type content. Instead, we transform the assets into audio and visual components. To transform web content into a form that is more similar to real TV program content, we could reduce the original web content or add other audio-visual information. At this time, when the information concerning the transformed content is changed from the original web content, the transmitted information differs from that of the original web content depending on how the system transforms the web content into TV-program-type content. In this paper, when there is little information change between the transformed content and the original web content, we regard the transmitted information as being high.
- **The dialogue among character agents**
  A user watches TV for various purposes; for example, sometimes to acquire information and sometimes just to enjoy watching TV. When a user wants to enjoy watching TV, animated character agents talk each other and they are better than only read-out. On the other hand, when a user wants to just acquire information, such a character agent generally is not needed. For the benefit of users who want to acquire different types of information for particular purposes, we also separate the audio and visual components into two types based on the user's purpose. We regard the way of communication between transformed content and user as how content appealing to a user.

3.3 Composition of transformed content

3.3.1 Audio component

The audio component is related to what information is communicated to users. In TV-program-type content, the audio component is read-out by using synthesized speech. It looks like the script of a real TV program but it is based on the text data in the web content. TV uses a range of broadcast types. Even within the news program category, there are different types - a headline news program involves news headlines presented by an anchorman, while morning and prime-time news shows present not only news reports but also commentary and background regarding the reports presented by newscasters. There are much greater differences between news and talk-show programs, e.g. a news program tends to simply report the bare facts, while a talk-show program uses dialogue to present news in an easily understood manner. In other words, these programs are scripted differently. If we simply want to catch up on the latest news, we can watch headline news, but if we want a detailed description of a news story, we can watch a news show, and if we want to watch TV for entertainment, we can watch a comedy talk show. As pointed out above, people sometimes watch TV to be informed and sometimes to be entertained. We describe two ways of transforming each component depending on these purposes.

Text read-out type

When a user simply wants to just acquire information from TV-program-type content, a system will communicate all the information in the web content to the user. In this case, the system transforms all the text in the web content into an audio component, so the transmitted information is high and it is without the dialogue among character agents.

Dialogue type

When the user wants to be entertained while receiving information, the system will transform web content into affinity content. We believe dialogue-based content increases user affinity. Most web content, however, is written in declarative sentences. For audio components, we therefore have to transform declarative-based web content into dialogue-based content. In this case, a system may communicate a summary of the information in the web content. If the entire web content is transformed into declarative content, the transformed content might become confusing. We transform web content into dialogue-based content by using keywords in the web content and eliminate unnecessary sentences in the web content. In this case, the transmitted information is low and there is a lot of dialogue among character agents.

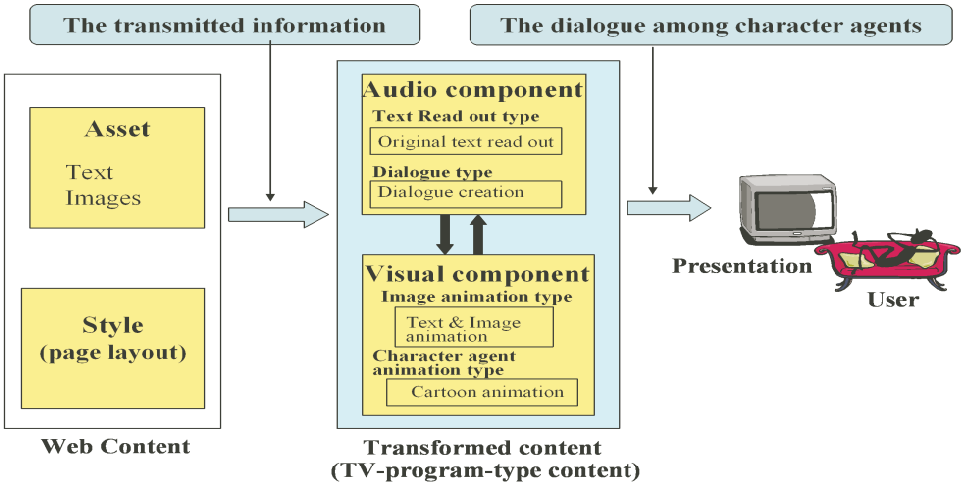

Figure 2: Architecture for transforming web content into TV-program-type content

3.3.2 *Visual component*

The visual component relates to how information is presented to users. In web content, visual assets include images, font size, and font color. In TV program content, visual assets relate to the direction of the content/program, and involve casting, characters, character action, studio set, camera work, and so on. In this way, the visual assets of web and TV program content are different. Users may want to get visual information intuitively, or in an entertaining fashion. In the former case, users may prefer a system that presents only image and text animation. In the latter case, the system can use character agent animation (cartoons) to present visually enjoyable content. Below, we describe the use of image animation type and character agent animation types to transform the visual component.

Image animation type

A user who wants only information might also want to receive visual information intuitively. We thus intuitively 'visualize' the information on a web page by using animation of the images, title, and text on the web page and synchronizing these animations with the synthesized speech. The resulting fusion content is simple TV-program-type content produced using Flash or SMIL, which enables users to easily and quickly acquire visual information from a web page just by watching TV.

Character agent animation type

If a user wants to be entertained while acquiring information from TV-program-type content, we 'visualize' the information by using character agents. When character agents are used to present web content, the user feels a stronger affinity for the content. In a character agent animation type presentation, the TV-program-type content consists of the character agent animation, telop, studio set, camera work, and set lighting. It is more similar to real TV program content because of the use of TVML.

3.4 Combination of audio and visual component

Our systems combine audio and visual component types, and transform web content into TV-program-type content. There are four possible combinations. However, we considered only three of these because the combination of dialogue type and image animation type is not realistic. We can create three types of content by changing the component types based on the user's preferences. Furthermore, by transforming the audio and visual components separately, we can also transform the content into a suitable form for mobile terminals just by changing only the visual content.

Table 1 shows the features of each pattern (with the names of our prototype systems in parentheses). Our prototype systems operate as follows: (1) u-Pav reads out text in web content and presents image animation along with text and keywords by ticker; (2) Web2TV reads out text in web content, automatically allocates the text in web content to several character agents, and presents images synchronized with the characters' speech; and (3) Web2Talkshow transforms summarized text in web content into a humorous character agent dialogue and presents character agent animation synchronized with the agents' dialogue. The u-Pav system consists of text read-out and image animation types, Web2TV consists of text read-out and character agent animation types, and Web2Talkshow consists of dialogue and character agent animation types. Figure 3 shows the relationship between the transmitted information and the dialogue among character agents. As shown, u-Pav provides high transmitted information, but without dialogue among character agents; on the other hand, Web2Talkshow provides low transmitted information, but allows much dialogue among character agents. Each user can select the best system according to his/her purpose.

4. TRANSFORMING DECLARATIVE CONTENT INTO DIALOGUE CONTENT

To transform an audio component into dialogue, we transform declarative sentences into dialogue sentences to create simpler and friendlier content. Although, the intended meaning of the original content may be unclear. However, we believe that if the content is transformed into dialogue based on keywords in the original web content, the intended meaning of the original should be preserved.

Table 1: The Features of Each Type Pattern

Combination name	Original web content	Dialogue	Keyword	Image	Image animation	Character agent animation
Readout & Image animation (u-Pav)	High	No	Yes	Yes	Yes	No
Readout & Character agent animation (Web2TV)	High	No	No	Yes	No	Yes
Dialogue & Character agent animation (Web2Talkshow)	Low	Yes	Yes	Yes	No	Yes

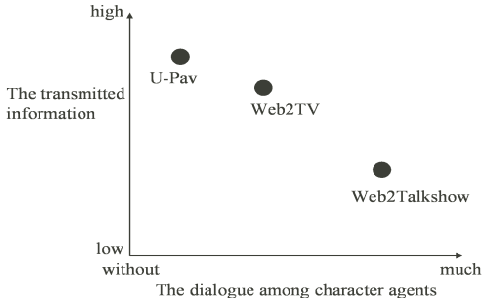

Figure 3: Relationship between the transmitted information and the dialogue among character agents

4.1 Extraction of keywords

Topic Structures

To extract the topic structures from a page, we use a topic-structure model. In the model, for a given page P, the topic structure t_i, $i \in \{1, \ldots, n\}$ is simply represented as a pair of *a subject term* s_i and a set C_i of *content terms*. C_i consists of multiple *content terms* c_{im}, $m \in \{1, \ldots, k\}$. Because a web page P may have more than one topic, s_i is associated with multiple c_{im}. That is, a topic structure t_i of a page P is represented as

$P = \{t_1, ..., t_i, ..., t_n\}$
$t_i = (s_i, C_i)$
$C_i = (c_{i1}, .., c_{im})$

- Subject terms
  We define the "subject degree" to determine whether a keyword has a high probability of being a subject term. The subject degree of word w_i is defined by its term frequency. The subject degree $sub(w_i)$ of keyword w_i within a given page P is defined as

  $$sub(w_i) = tf(w_i) \times weight(w_o) > \alpha$$

  where $tf(w_i)$ denotes the term frequency of w_i on the corresponding page, $weight(w_i)$ denotes the weight of w_i, and α denotes a threshold. Regarding the weights, we weight proper nouns, numbers, and common nouns more heavily in that order. If all nouns and proper nouns have the same weight, numbers and numerical classifiers will have higher weights. In our experiments, the following weights are assigned to different types of nouns: "proper nouns" as 3.0, "numbers" as 0.1, "numerical classifier" as 0.1, "general nouns" as 1.0, and "other nouns" as 0.9. The word vector is equal to the word frequency multiplied by the word weight.

- Content terms
  The content terms in a given page are intuitively those with a high co-occurrence relationship with a specific subject term in the page. We use co-occurring words in the previous pages to extract content terms. To determine the content terms, we prepare a matrix for the term co-occurrence data in advance. Suppose that the content degree $con(w_i)$ of keyword w_i within P is defined as the sum of undirected term co-occurrence rates with the subject terms of w_i.

  $$con(w_i) = cooc(w_i, S) > \beta$$

  where S is the subject-term set and $cooc(w_i, S)$ is the co-occurrence rate between w_i and S.

4.2 Transforming Basic Dialogue

Our transformation method is only a partially automatic operation because there are still difficulties in completely automating the operation. We write a dialogue framework in XML as a pre-scenario, and create a scenario based on the original web content's story-flow and topic structure of the original web content along with the pre-scenario. We write many different dialogue patterns in the pre-scenario, and the system chooses a dialogue pattern based on the structure of the sentences in the original web content. We divide the sentences into two types, i.e. sentences that contain/do not contain the subject term, with the transformation into dialogue being based on sentence type.

4.2.1 A Sentence Including the Subject Term(s)

In this case, we transform a sentence into a question-and-answer based dialogue. Subject terms and content terms are nouns only, and do not become predicates. Thus, we focus on a subject and an object in a sentence as follows:

- Subject term is a subject in a sentence

There are two cases based on a dependency analysis of the sentence.

- **Content terms dependent on the subject**
  In this case, the subject term and content terms are strongly related. We transform the sentence into a question-and-answer dialogue in terms of the relationship between the subject term and content terms.
  For example, in the sentence, "Ichiro of the Seattle Mariners set a world record today at Safeco Field." the subject term is "Ichiro" and the content term is "Mariners". The system transforms the sentence into a dialogue as follows:
  A: Who is Ichiro?
  B: I know. Ichiro is one of the Mariners.
  A: That's right! He set a world record today at

Safeco Field.
Thus, to create a dialogue, the system uses the subject term(s) and the content term(s) as the question-answer. After the first question-answer, the system changes the subject to a pronoun, e.g., "Ichiro of the Seattle Mariners" becomes "he". When the subject term is a person's name, the system uses who-type questions. But when the subject term is a different type of noun, it uses 'what' questions.

- **Content terms independent of the subject**
In this case, the subject term is not strongly related to the content terms. We transform the sentence into a dialogue in which the answer is a verb or object that co-occurs with the subject terms in the sentence. For example, in the sentence, "Ichiro played for the Mariners against the Rangers at Safeco.", the subject term is "Ichiro" and the content term is "Safeco". The system transforms this into a dialogue as follows:
A: What did Ichiro do?
B: I know. He played.
A: What did he play?
B: He played for the Mariners against the Rangers at Safeco.
A: That's right.

- Subject term is an object in a sentence
In this case, the answer in the dialogue is based on a term that depends on the subject term. For example, in the sentence, "He extended the world record to 262.", the subject term is "record" and it depends on the object of "extend". The resulting dialogue goes as follows:
A: What did he do?
B: He extended the world record.
A: Is that right?
B: That's right.

4.2.2 A Sentence Without a Subject Term

When a sentence includes a date or place, it is likely to be important to the meaning of the sentence. In this case the system transforms the sentence into 'when' or 'where' questions.
For example, in the sentence, "The Mariners lost a game on October 30.", the subject term is again "Ichiro", and the resulting dialogue is as follows:
A: When did the Mariners lose a game?
B: On October 30.
A: You are smart!!
When a sentence does not include a date or place, we presume the sentence does not contain important terms. The system also transforms a sentence into yes/no questions or tag questions.
For example, in the sentence, "George Sisler held the 84-year-old single-season hit record.", the subject term is still "Ichiro". The dialogue then goes as follows:
A: George Sisler held the single-season record, didn't he?
B: Yes. That's right
A: Humm..

4.3 Transforming Dialogue with Humor

We can also transform declarative sentences into humorous dialogue. We believe that humor is the easiest way of ensuring that users of any age understand the content. However, we have to pay attention to the type of original content. Some types of content are not suitable for transformation into humorous content, for example, news concerning serious matters or accidents. In the real world, humorous dialogue uses exaggeration, deliberate mistakes or misunderstandings, or surprise twists; i.e., a humorous dialogue is often based on taking a strange or unexpected point of view of a common situation. We believe we can use these methods to transform content into humorous dialogue.
Mistakes and misunderstandings
We transform content into dialogue based on a topic structure consisting of a subject term and content terms. Content terms co-occur with the subject term; i.e., content terms are terms ordinarily used with the subject term. We believe that if the system deliberately uses mistaken topic structure sets consisting of incorrect content term(s) and a subject term, to transform dialogue, the system can transform sentence into humorous dialogue based on the mistakes. To create this type of dialogue, we extract an incorrect topic structure from a different topic graph of an entire page. If a page consists of multiple topics, we extract incorrect content terms that are indirectly connected to the subject. For example, when the correct topic structures t1, t2, and t3 for an entire page are
$t1 = (s_1, \{c_1, c_2\}) = (Ichiro, \{Mariners, hit\})$
$t2 = (s_2, \{c_3\}) = (Rangers, \{Texas\})$,
we create an incorrect topic structure $it1$ as follows:
$it1 = (s_1, \{c_5\}) = (Ichiro, Texas)$
In this case, the question type is
A: Who is Ichiro?
B: Ichiro is Texas.
A: What? Ichiro is Texas?
B: In my town, everybody says so.
A: Oh no! Ichiro is a Mariner who set the world record for hits.

Exaggeration
The first step in creating dialogue based on exaggeration is to use bigger numbers. When a sentence includes numbers, the system increases the numbers by a substantial factor. For example, if the sentence is "Today, I picked up $1", the exaggerated dialogue becomes:
A: Today, I picked up $1000.
B: Wow! $1000?
A: Oh, I made a mistake. I picked up only $1.

4.3.1 Pre-scenario

We create dialogue frameworks in a pre-scenario file in XML. The pre-scenario consists of structure tags, content tags, and direction tags. Table 2 shows the pre-scenario structure and content tags.
For the above example, we would write the dialogue framework as follows:
< *question type* = "1 − 2" *key* = "*who*" >
< *line chara* = 1 > $question1 ? < *line* >
< *line chara* = 2 > I know, $answer1 < *line* >
< *line chara* = 1 > That's right! $sentence1.< *line* >
< *question* >

```
<?xml version="1.0" encoding="EUC" ?>
<initialize>
        <set up type="1">
</initialize>
<Intro>
        <dintro type="1">
        <line chara="1"> Hello, I'm Bob.</line>
        <line chara="2"> Hi!!, I'm Mary.</line>
        <line chara="1"> Do you know today's topic?</line>
        <line chara="2"> I know we will talk about $theme, today.</line>
              ⋮

</Intro>
<Dialogue>
        <question type="1-1" key="what">
              <LookAtCamera>
              <line chara="1"> Do you know, $sentence1?</line>
              <nod>
              <line chara="2"> Of course!! $mis-answer!! </line>
            <line chara="1">Waoo!! You are foolish!! $answer. </line>
                ⋮
        </question>
        <question tyoe="1-2" key="who">
              <line chara="1"> $sentence1?</line>

        </question>
         ⋮
</Dialogue>
<Conclusion>
         ⋮
</Conclusion>
```

Figure 4: Example of a Pre-scenario

Table 2: Pre-scenario Structure and Content Tags

Structure Tags	
Initialize	The pre-processing is enclosed in this tag.
Intro	The introduction part is enclosed in this tag. It consists of a greeting and the part where the theme of the original web content is described in the body.
Dialogue	The body part is enclosed in this tag. This part is the main part that is transformed into TV-program-type content.
Conclusion	The conclusion is enclosed in this tag. This part consists of a farewell and the final laugh line, which is a joke based on the theme of the original web page that is created using a joke dictionary.
Content Tags	
line	character agent' speech line. This attribute is a character that specifies which character speaks in this line.
nod	character agent's nod line.
surprise	character agent's surprise line.
question	This specifies the question-type framework. By using the "type" attribute, users can choose a specific type of question from a range of types.
exaggeration	This specifies the exaggeration framework. By using the "type" attribute, users can choose a specific type of exaggeration from several types.

In the 'type="1-2"', the number before the hyphen "1" represents the type of sentence relating to the topic structure. The number after the hyphen, "2", represents the number of variations of the question type. We can create different frameworks by creating variations of each sentence. This example is variation 2 of sentence number 1, which includes subject term(s) and content terms dependent on the subject. We input transformed questions and answers into each variation. Figure 4 shows an example of a pre-scenario.

5. PROTOTYPE SYSTEMS

We developed three types of prototype systems. These are called u-Pav, Web2TV, and Web2Talkshow.

5.1 u-Pav

The u-Pav (Ubiquitous Passive Viewer) system is based on using the text read-out type of the audio component and the image animation type of the visual component. We developed u-Pav for two purposes: (1) to adapt fusion content

Figure 5: Display Image of u-Pav

Figure 6: Display Image of Web2TV

to a ubiquitous environment, and (2) to communicate the entire content of selected web pages to users accurately and intuitively. The audio component of u-Pav is text, which is articulated using synthesized speech. For the visual component, the title and lines are shown through a ticker, and keywords and images are animated. The program synchronizes the tickers, the animations, and the speech. Figure 5 shows a display image from u-Pav. U-Pav can be displayed on a mobile phone screen simply by changing the visual component. The system was designed for use in a business environment. We developed u-Pav using Flash because Flash content can be displayed on mobile phones in Japan.

5.2 Web2TV

Our Web2TV prototype system presents audio components using text read-out types and visual components using character agent animation types. Web2TV looks like a headline news program. Character agents are used to read-out the web content and the system presents images synchronized with the character agent reading out the text. The audio component consists of the character agents' lines,

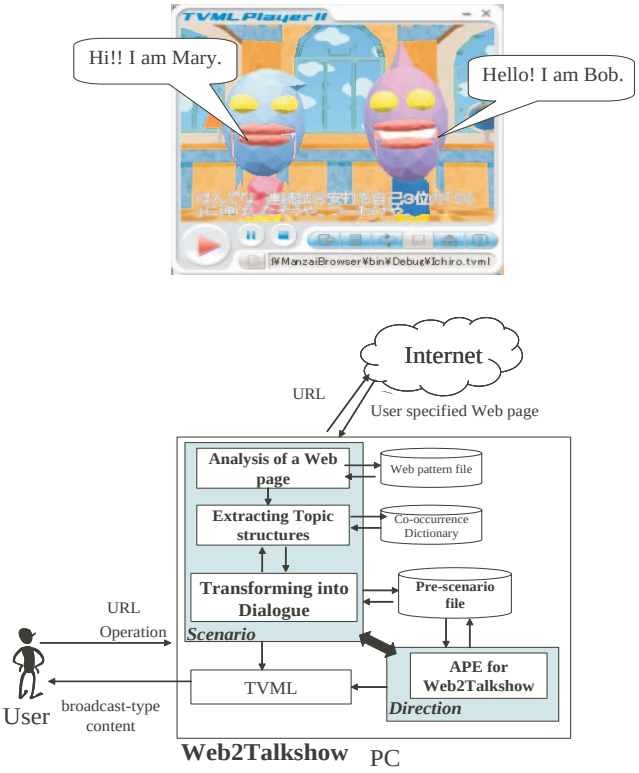

Figure 7: Display Image and system architecture of Web2Talkshow

which are the entire web page text. The visual component consists of the camera work, lighting, studio set, and the character agents and their actions. The character agent actions are not intense. With Web2TV, we can easily combine news programs with news web content, and we can create news-program-type fusion content. We have also developed a Web2TV mobile phone version. Figure 6 shows a Web2TV display image.

5.3 Web2Talkshow

Web2Talkshow uses the dialogue type for audio component and the character agent animation type for the visual component(see Figure 7). In our prototype system, we transformed declarative sentences from the web content into humorous dialogue. In Japan, there is a traditional form of comedy called "manzai". "Manzai" typically consists of two or three comedians participating in a humorous dialogue, rather like American "stand-up comedy", or Chinese "xiang sheng". In the case of two people, one is the "straight man", the other is the "fool". We use this "manzai (or "stand-up comedy" or "xiang sheng")" style in Web2Talkshow. The audio component consists of the character agents' dialogue with humor lines. The visual component is the same as in Web2TV. With Web2Talkshow, however, the character agents talk to each other and their actions depend on their conversation, as in real life. The direction required for individual dialogues is time-consuming, however, so we create numerous visual components depending on the lines using XML. The direction tags are transformed into animated character agents using APE[10] for Web2Talkshow. Users can develop other APE and XML tags to create other types of animated character agents. The benefit of using the APE for Web2Talkshow is that the system can easily transform either the same visual component and different audio component, or different visual component and the same audio component into many types of TV-program-type content. By using Web2Talkshow, we can easily combine a talk-show program with transformed content, and create easily understood fusion content.

6. EXPERIMENT OF USER EVALUATION

We did three types of user evaluation: experiment 1 was a comparison of Web2TV and Web2Talkshow, experiment 2 was to determine user attitudes towards Web2Talkshow for various age groups, and experiment 3 was a comparison of u-Pav and Web2Talkshow.

6.1 Experiment1: Comparison of Web2TV and Web2Talkshow

We evaluated the usefulness of our transformed content using Web2Talkshow and Web2TV. We did a user evaluation experiment with 120 subjects. There were an equal number of male and female subjects divided into three groups of 20 according to the following age groups: 20-34 years old, 35-49 years old, and 50 years old or more. They were not computer professionals. First, we showed the same news page using Web2TV and Web2Talkshow. Next, the subjects transformed their four favorite news pages into TV-program-type content using Web2Talkshow. Our questions were as follows:

1. Which did you understand better, Web2TV or Web2Talkshow?
2. Did you feel the keywords were appropriate?
3. Did you understand the character agent dialogue?
4. Did Web2Talkshow convey the web content accurately?
5. Is Web2Talkshow more affinity content than Web2TV?

The results of the experiment are shown in Figure 8 The results of the questions did not depend on the age of the subjects. In the results for Q1, 31% of the subjects preferred Web2Talkshow and 52% preferred Web2TV, i.e. almost half of the subjects liked Web2TV more than Web2Talkshow. For Q2, 75% of the subjects said the keywords were appropriate, indicating that our keyword extraction methods were suitable for extracting keywords from web pages. For Q3, 80% of subjects understood the character agent dialogue, which meant the system did a good job of transforming web content into dialogue content. In the results of Q4, 65% of subjects said the system conveyed web content accurately. However, comparing the Q3 and Q4 results, we see that at least 15% of the subjects thought the system produced good dialogues, but did not convey the web content accurately. For Q5, 60% of the subjects said character agent animation type content created stronger affinity. Thus, in our future work we will investigate methods for accurately transforming content into dialogue. Overall, the results showed

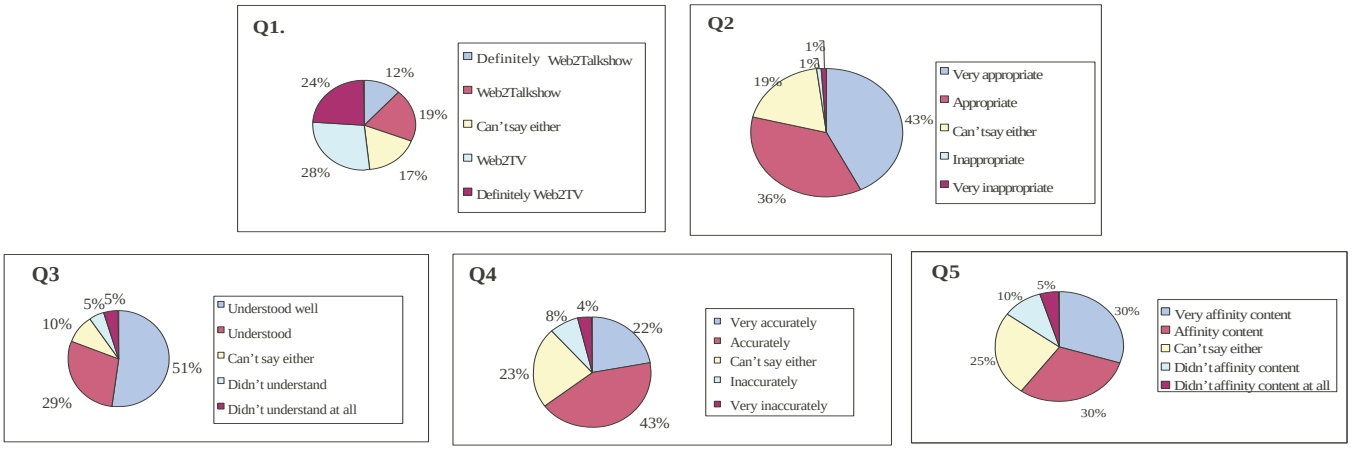

Figure 8: Results of User Evaluation1

that Web2TV and Web2Talkshow were both useful, and our technique of separating audio and visual components was suitable for transforming web content into TV-program-type content. Typical favorable comments included

- Some users said that the transforming web content into TV-program-type content concept was very interesting, and by making them digital divide people can get information from the Internet is very good.
- Users said it was easy to understand the original web content after transformation into dialogue based on web page keywords.

Typical unfavorable comments were

- The accent of the synthesized speech was strange and differed from a real speech accent.
- When the dialogue was long, it was a little difficult to understand the original web content.

Based on these comments, we need to improve the interval between dialogues and the dialogue length.

6.2 Experiment2: User attitudes to Web2Talkshow based on age

We measured user attitudes towards Web2Talkshow according to age through an experiment with 172 subjects. Figure 9 shows the distribution of the subjects by age. While 10% of the subjects was computer researchers or specialists, the remaining 90% had no special experience in this area.

We performed transformations of two kinds of news page into TV-program-type content based on traditional manzai comedy. We had the subjects complete a questionnaire containing the following questions:

1. How interesting was it to use Web2Talkshow?
2. Did transforming the news page into a manzai format make it easier for you to use?
3. Does Web2Talkshow make it easier to understand what is written on web pages?

We categorized the subjects who were less than 20 years old as children, subjects from 20 to 59 years old as adults, and subjects older than 59 as elderly. The results of the experiment are shown in Figure 9. Before the evaluation, we expected Web2Talkshow to be especially appealing to children, but the results were only weakly related to the age of subjects. It is noteworthy, though, that half of the subjects in the child group found Web2Talkshow to be very interesting (the response to Q1), suggesting that Web2Talkshow makes a more positive impression on younger users.

6.3 Experiment3: Comparison of u-Pav and Web2Talkshow

In the third experiment, we had 50 subjects compare u-Pav and Web2Talkshow. The subjects ranged from 20 to 50 years in age. We transformed the same sports news in a web site into u-Pav and Web2Talkshow. About 60% of subjects said u-Pav was more interesting than Web2Talkshow. However, the younger subjects found Web2Talkshow more amusing. This suggests adults prefer u-Pav to Web2Talkshow, which is encouraging since we developed u-Pav for a business environment. In this experiment, we only used the PC version of u-Pav, but half of the subjects said they thought u-Pav would be convenient on a mobile phone. Because it used image animation and ticker animation, most subjects felt they would be able to understand a news report by just watching u-Pav. Most subjects also said Web2Talkshow was better for children than u-Pav. Only adult subjects participated in this experiment, so we plan to repeat the comparison with children and older people among the subjects.

6.4 Discussion

Comparing the results of these three experiments, we found those of experiment 2 regarding Web2Talkshow were the most encouraging. However, the experimental settings differed significantly. In experiments 1 and 3, all of the subjects were adults (over 20 years old) and they each watched Web2Talkshow on a PC. In contrast, the adult subjects in experiment 2 were parents or grandparents who gathered with their children to watch Web2Talkshow together. Thus, these subjects probably felt more relaxed during the experiment. These different situations might have affected the results. For example, when users want to relax while watching TV, they are more likely to feel that humorous dialogue is more comfortable and more interesting. When they want

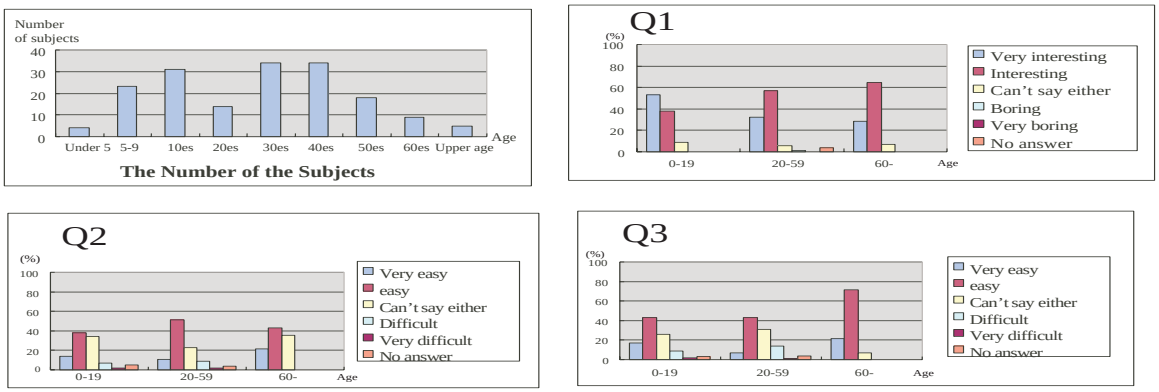

Figure 9: Results of User Evaluation2

to watch TV to obtain information, they generally prefer to get the information directly, such as through Web2TV or u-Pav. How user attitudes towards our proposed systems differ depending on the situation is an issue we are now planning to investigate.

7. CONCLUSIONS

We have developed ways to automatically transform web content into TV-program-type content as the first step towards media fusion. Our transformation systems are based on creating audio and visual components. Here, we have described the use of text read-out and dialogue techniques for transforming the audio component, and image animation and character agent animation types for the visual component. By combining these techniques, we can transform web content into various types of TV program content, and we can fuse this with various broadcast programs. In this paper, we have also explained how declarative content can be transformed into dialogue content using the topic structure. Our evaluations of three prototype systems u-Pav, Web2TV, and Web2Talkshow have shown the usefulness of our approach. In our future work, we plan to work on techniques for refining the fusion of transformed content with TV program content.

8. ACKNOWLEDGMENTS

Research on Web2Talkshow was carried out in collaboration with NHK Science & Technical Research Laboratories, and on u-Pav with Nomura Research Institute. We thank Dr. Hayashi, Dr. Douke, and Dr. Hamaguchi of NHK Science & Technical Research Laboratories, and Dr. Yokozawa, Mr. Hamabe, and Mr. Uwada of Nomura Research Institute for helpful discussions and comments.

9. REFERENCES

[1] *Feeddemon site homepage.* http://www.bradsoft.com/feeddemon/index.asp.

[2] *Sharpreader homepage.* http://www.sharpreader.net/.

[3] *TVML site homepage.* http://www.strl.nhk.or.jp/TVML/.

[4] *W3C Authoring Techniques for Device Independence.* http://www.w3.org/TR/2004/NOTE-di-atdi-20040218.html.

[5] *WebTV site homepage.* http://www.webTV.com/.

[6] H.Miyamori and K.Tanaka. Webified video: media conversion from tv programs to web content for cross-media information integration. In *DEXA2005, LNCS3588, Springer-Verlag*, pages 176–185, August 2005.

[7] H.Nonogaki and H.Ueda. Friend21 project: A construction of 21st century human interfac. In *International Conference on Human Factors in Computing Systems(CHI'91)*, pages 407–414, April 1991.

[8] J.Zimmerman, N.Dimitrova, L.Agnihotri, A.Janevski, and L.Nikolovsa. Myinfo: a personal news interface. In *International Conference on Human Factors in Computing Systems(CHI'03)*, pages 898–899, April 2003.

[9] K.Tanaka, A.Nadamoto, M.Kusahara, T.Hattori, H.Kondo, and K.Sumiya. Back to the tv: Informationvisualization interfaces based ontv-program metaphors. In *Proceedings of IEEE International Conference on Multimedia & Expo (ICME2000)*, pages 1229–1232, August 2000.

[10] M.Hayashi, M.Douke, and N.Hamaguchi. Automatic tv program production with apes. In *The 2nd International Conference on Creating, Connecting and Collaborating through Computer(C5), IEEE Press*, pages 20–25, January 2004.

[11] M.Ishizaki and Y.Den. *Computation and Language Volume3:Discourse and Dialogue.* University of Tokyo Pressy, 2001.

[12] Q.Ma, A.Nadamoto, and K.Tanaka. Complementary information retrieval for cross-media news contents. In *Proceedings of ACM MMDB 2004*, pages 45–54, November 2004.

[13] Q.Ma and K.Tanaka. Webtelop: Dynamic tv-content augmentation by using web pages. In *Proceedings of IEEE International Conference on Multimedia & Expo (ICME2003) Vol.2*, pages 173–176, July 2003.

[14] T.Kamba, H.Sakagami, and Y.Kosekia. Automatic personalization on push news services. In *Proceedings of W3C Push Workshop*, September 1997.

[15] T.Yamaguchi, I.Hosomi, and T.Miyashita. Webstage: An active media enhanced world wide web browser. In *International Conference on Human Factors in Computing Systems(CHI'97)*, pages 391–398, April 1997.