We can now obtain textual information that describes program content in the form of an electronic program guide (EPG) or closed captions in digital broadcasting. A textual program summary in the EPG provides a simple description of a program’s content, while closed captions offer detailed explanations of video content. This article discusses technologies for recognizing program content through linguistic analysis of the textual information.

**Video Abstraction Based on the Sentence Similarity**

Video abstraction is a technique that can be used to efficiently browse to a desired program from a massive amount of video data. Abstraction to create a program introduction requires the selection of scenes based on the semantic content of the program. We proposed a method that uses similarity between sentences in the textual program summary and in the closed captions for producing the abstracted program video. The method identifies sentences in the closed captions that are most similar to sentences in the summary, and links the corresponding video segments (Figure 1).

**Automatic Generation of Human Relationship Chart**

We consider a graphical representation of a textual program summary as a way of assisting users to grasp the program’s content. For example, in an article introducing movies in a magazine, a human relationship chart often appears along with the main article to help with comprehending the content. We developed a technique to automatically generate a human relationship chart using word characteristics and sentence structure in the program summary, which identifies words that express people and their relationships. When combined with face recognition, facial images obtained from the video can also be displayed on the relationship chart (Figure 2).

**Identification of Principal Objects and Actions**

Identification of principal objects is important for searching for a specific scene in a program. It is however difficult to identify a principal object only using conventional video analysis because many types of objects appear in a typical video segment. We proposed a technique to identify a principal object using the linguistic characteristics of closed captions when the principal object appears in the video segment. We also developed a method to detect actions of the principal object based on the analysis of the words that appear with the object (Figure 3).
The previous articles in this series introduced the concept of metadata and the various video/image analysis and text processing methods that can be used to generate metadata automatically. There are a number of technologies available these days for generating useful information for metadata applications. However, future advances will require appropriate combinations of these technologies in order to extract deep-level semantic data from video content, and combining such technologies is not an easy task. While many research institutions have independently developed outstanding information processing technologies, their use of different development environments poses obstacles to combining these technologies for greater effect.

This is the background for our proposing the metadata production framework (MPF) (Figure 1). The purpose of this framework is to provide a common foundation to produce metadata. The MPF includes a definition of common data format and rules for exchanging the data, with the intention that research and development on information processing being carried out in various institutions can be easily incorporated into new systems. Its modular design protects proprietary technology, since the internal mechanism cannot be seen from outside the module. The MPF incorporates independent technologies in a coordinated metadata production process. The framework also works in a distributed environment. The MPF standard is available at STRL’s homepage (http://www.nhk.or.jp/strl/mpf/index.htm). Besides the specifications document, other related software, such as the common base metadata editor (Figure 2), can be downloaded. A sample modular program included in the software allows a user to immediately try out the framework to see how it works or modify it by embedding his/her own processing. We invite you to try the MPF.
STRL researchers are imagining a new TV viewing style, called CurioView. For this new style of viewing to be possible, we need functions that use metadata for program-related information retrieval.

**What Is CurioView?**

When we watch TV, we usually select programs from a list including live and recorded programs. The program currently being watched often reflects our interests at that particular time. If additional content can be provided, such as related video footage or information, when we decide that we are interested in more details about the program, it would help to satisfy our desire for knowledge and naturally expand our scope of interest (Figure 1). This TV viewing style is called CurioView.

**Mechanism and Prototype**

CurioView requires the TV to have an ability to retrieve related information automatically by referring to metadata that describes the program content. By combining the metadata extraction techniques that have been featured in this series of articles, such as video analysis and language processing, we have developed a technology for metadata indexing based on programs and scenes. This technology is used to perform data matching and relevance evaluations (see Figure 2). With this technology, the CurioView prototype system can automatically retrieve video content and information related to the program currently being viewed (Figure 3). Furthermore, we are improving baseball video search function by defining various concepts regarding baseball games and incorporating a data retrieval technology that searches for highlight scenes that reflect human sense, such as scenes where there is the chance of a come-from-behind win.

Our future work will enhance the metadata extraction accuracy and the information retrieval and display methods by exploiting semantic relationships found in program content.

Note)

CurioView is derived from the words, curiosity and view.