An XML Based Multimedia Data Acquisition and Retrieval With Wearable Computers

Yoshihiro Ohmori, Kazushige Ouchi, Masakazu Hattori, and Miwako Doi
TOSHIBA Corporate Research and Development Center
1, Komukai Toshiba-Cho, Saiwai-Ku, Kawasaki 212-8582, JAPAN
+81-44-549-2286
{yoshihiro.ohmori, kazushige.uchi, masakazu.hattori, miwako.doii}@toshiba.co.jp

Abstract

We developed an XML based multimedia data acquisition and inquiry system for wearable computers. This system can record multimedia data in mobile environments, stores recorded data into an XML native database in organized XML formats, and provides an easy access method to the stored data. An XML based general framework within the system supports users to browse search results in appropriate forms.

Key words: context-awareness, multimedia database, XML, indexing, wearable

1. Introduction

Context-awareness is important for wearable computers [1]. The touring machine case study uses the context-aware knowledge for how information is viewed [2]. On the other hand, the field study prototyping system uses context-aware knowledge for how data is recorded [3]. We consider that context-awareness can provide more useful functions when used in both recording and viewing.

This paper describes a multimedia data acquisition and inquiry system in XML (eXtensible Markup Language [7]) with a wearable computer. This system can index and organize multimedia data by recognizing spoken key words and attaching voice annotations to the pictures taken by the system.

At the beginning of this paper, problems of existing wearable multimedia recording systems are discussed. Next, the XML based multimedia data acquisition and inquiry system is introduced. And next we show an example to gather multimedia data, and ensure the efficiency of this system. Finally we conclude this paper and describe future works.

2. Problems of Existing Wearable Multimedia Recording Systems

Recent evolution of technology enables consumer products to record multimedia data. For example, digital still camera can record not only pictures but also voice, movie, and handwritten notes. However, such voice memo or handwritten notes are not be used for picture retrieval.

There are a lot of researches about wearable computers, and they gather huge size of multimedia data. But recorded multimedia data can be used only within the system. So users can only browse recorded data in static forms that the system provide.

These existing recording systems are designed for gathering multimedia data and they do not help people to recall the relationships among the gathered data and do not support access methods to the gathered data. In other words, these existing systems do not provide method to change how to browse recorded data.

We had developed trial systems to link recording and retrieving multimedia data [5] [6]. The system organizes recorded multimedia data, and provides access method to the data for database engineers [5]. The authoring system records multimedia data in wearable environment, and stores them in HTML format [6]. In the system, users can browse recorded data in organized forms in HTML. But users have to re-program the system to generate other forms of HTML to get other views of recorded data. It will cost very expensive.
To use recorded data in more general syntax, storing recorded data in XML is a pretty good solution. Especially, MPEG-7 (Moving Pictures Experts Group phase 7) is supposed to be suitable to describe attributes about multimedia data. However, MPEG7 is on the way of standardization, so it is reasonable to adopt parts of MPEG7 standard toward well-describable attributes. Once data is described in XML, it is relatively easy to convert XML family languages like XHTML, SMIL (Synchronized Multimedia Integration Language [4]), and so on.

People use several authoring systems in order to make presentation materials, i.e., reports or Web Pages in HTML, streaming presentations in SMIL. In the exhibition tour for example, people take a lot of pictures with a digital camera and hand-written notes. Then they prepare a report assembling a lot of digital pictures and notes at their office or home. In this case, the vivid information is lost and it is bothersome to assemble the unorganized multimedia data.

The purpose of SMIL is to play multimedia materials synchronously. The available implementations of them only realize synchronous play of multimedia materials, and do not realize information retrieval. The purpose of MPEG7 is to describe multimedia materials in standardized forms. But the standard does not fix how to search multimedia data.

Target users for authoring multimedia presentation materials in SMIL or MPEG7 are professionals: Internet broadcasters, digital broadcasters, image processing researchers, and so on. They may be accustomed to retrieve data from independently recorded multimedia data. But we think that information retrieval is the most critical function in order to recall the relationships among the gathered data.

To overcome these problems, we introduce an XML based multimedia data acquisition and retrieval system for wearable computers.

3. System Overview

Figure 1 shows an overview of the system. The system is composed of three components, i.e., a multimedia data recorder, a multimedia database (MMDB), and a viewer. A multimedia data recorder records multimedia data, and recorded data are described in XML format. A multimedia database stores multimedia data written in XML, and search data by queries written in XML. A viewer issues queries to the MMDB, and displays search results by transforming them into HTML. In this way, all the attribute information is exchanged as XML instances among each component. We will describe each component of the system in detail below.

![Figure 1. System Overview](image)
3.1. Multimedia Data Recorder

The multimedia data recorder runs on a wearable computer shown in Figure 2. The wearable computer consists of a small computer (TOSHIBA Libretto ff 1100v), a hat-mounted camera (an option camera of Libretto ff 1100v), a microphone, and a remote switch.

At one time, the multimedia data recorder records pictures attaching voice memos, date of record, and keywords that are recognized by a speech recognition module (TOSHIBA speech recognition module). The remote switch sends action commands to the computer; i.e. capturing a picture, recording voice memo, and recognizing speech key words. The multimedia organization module on the small computer links these data synchronously, and organizes them [5] [6]. In the recorder, recorded data are described in XML format.

![Appearance of a Wearable Computer](image1)

**Figure 2. Appearance of a Wearable Computer**

3.2. Multimedia Database

The multimedia database runs on desktop / laptop computers, and receive recorded data from the recorder via a wireless LAN. The Multimedia Database is composed of an XML native database and a file system. The XML native database stores textual information of data, i.e. dates of record, links between data, and file paths of multimedia contents in URL. The XML database does not store multimedia contents in itself, and only holds these textual information. Multimedia contents such as pictures and voice memo are stored into a file system as Bitmap image files and wave audio format files.

In this system, we use TOSHIBA Knowledge Factory XML Database (KF) as XML database, which can directly store XML instances and search XML fragment by issuing queries written in XML. As contrasted with traditional relational, KF is suitable for searching complexly structured data.

3.3. Viewer

The viewer is a client software of the multimedia database described above, and runs on desktop / laptop computers. The viewer inputs user instruction with GUI, and displays search results in appropriate forms to users.

Figure 3 shows an appearance of the viewer. The viewer is composed of three panes; left upper pane shows category of prepared search operations, left lower pane is for inputting free keywords, and right pane shows search results.

Selecting a category in left upper pane of Figure 3 cause issuing a query written in XML onto the multimedia database described before. Then the viewer transform search results to XML based HTML (XHTML) according to a style sheet. Style sheets are written in XSLT (Extensible Stylesheet Language Transformations), and prepared in advance.

The category helps users to select frequently executed search operations. Although professional users can generate queries, amateur users are not accustomed to making queries. Therefore, it is useful to prepare search operations in advance like this category. Furthermore, the viewer can categorize prepared search operations by displaying them in tree forms in the category pane shown in Figure 3.

One category is composed of a label for displaying on the viewer, a query, and style sheets for displaying search results. Because the viewer reads a category file written in XML, users can easily customize the viewer by preparing a custom category file.
4. Example

To examine the efficiency of this system, we adapt this system to actual data sets. We had an opportunity to describe a panel at a conference, and planned to record information about attendees. Our goal is to record who is interested in which topic of the panel, because these information is likely to be forgotten.

By using multimedia data recorder, we recorded pictures of faces, names by voice memos, occupations and interests by using speech recognition, and record dates while describing the panel. Instructions with remote switch enabled to explain the panel without interruptions and stress. Finally we could get 33 data sets.

List 1 shows an example of one data set written in XML. <MediaInformation> tag means a data element, which contain one data set, <TitleImage> tag means a picture of an attendee, <Sound> tag means a file path of a voice memo, <RecordDate> tag means date of record,

<Occupation> tag means an occupation of the attendee, and <Interest> tag means an interest of an attendee.

Recorded data are immediately sent to MMDB running on a laptop computer beside of the panel. After the conference, we browsed data to make a report about the conference by using the viewer.

List 1. Example of a Recorded Data
List 2 shows an example of a query to retrieve data sets about attendees who are interested in "groupware". The query means to search data sets that contain a "groupware" key word in <interest> tag.

KF adopts SQL like syntax for query expression. <kf:from> tag specifies which data have to match, and <kf:select> tag specifies output format. In this example, target data for search is set to "uix://root" by a "path" attribute in the <kf:from> tag, and KF searches data sets that has the same structure within the <kf:from> tag. Data elements starting with "$" are variables, and data elements without "$" are restrictions. When <interest> element is matched in the database, variables in other elements are bound, and output them according to the format within the <kf:select> tag. <results> tag is used for describing repetition of results.

```
<results>
    <kf:query xmlns:kf="">
        <kf:select>
            <result>
                <Title>$occupation</Title>
                <TitleImage>$titleimage</TitleImage>
                <Sound>$sound</Sound>
                <RecordDate>$recorddate</RecordDate>
                <Occupation>$occupation</Occupation>
                <interest>groupware</interest>
            </result>
        </kf:select>
        <kf:from path="uix://root">
            <wauth>
                <MediaInformation>
                    <TitleImage>$titleimage</TitleImage>
                    <Sound>$sound</Sound>
                    <RecordDate>$recorddate</RecordDate>
                    <Occupation>$occupation</Occupation>
                    <interest>groupware</interest>
                </MediaInformation>
            </wauth>
        </kf:from>
    </kf:query>
</results>
```

List 2. Example of a query

List 3 shows an example of a result searched by the MMDB. Each matched results are listed in the <results> tag in the query shown in List 2. The viewer transforms this result to XHTML according to an XSLT style sheet, and displays it to users like shown in Figure 3.

```
<results>
    <result>
        <Title>company</Title>
        <TitleImage>
            New_006_HTML/image000.bmp
        </TitleImage>
        <Sound>
            New_006_HTML/sound0.wav
        </Sound>
        <RecordDate>2000/03/14T12:15:08</RecordDate>
        <Occupation>company</Occupation>
        <interest>groupware</interest>
    </result>
    <result>
        <Title>university</Title>
        <TitleImage>
            New_030_HTML/image000.bmp
        </TitleImage>
        <Sound>
            New_030_HTML/sound0.wav
        </Sound>
        <RecordDate>2000/03/14T16:31:21</RecordDate>
        <Occupation>university</Occupation>
        <interest>groupware</interest>
    </result>
    ...
</results>
```

List 3. Example of a Search Result

Furthermore the viewer can generate variety types of appearance to display search results by changing style sheets. Figure 4 shows that users can browse two types of search results from the same result. Right window dump shows a list style, and left window dump shows a calendar style. In response to users needs, this system can provide appropriate forms of search results.

Through the example study, we could ensure that this system can record multimedia data without stress, and help users to analyze recorded data without editing. For database developers XML based framework within the system enables easy customization.
5. Conclusions and Future Works

In this paper, we developed an XML based multimedia data acquisition and retrieval system for wearable computers. This system can record multimedia data in mobile environments, and help users to retrieve recorded data in appropriate forms. Through an example study, we ensured that this system can record multimedia data without stress, and that XML based framework within the system supports users to browse search results in appropriate forms.

We think the developed system is also useful daily reports and Web pages generation by automatically organizing multimedia data. We are planning to add other recognition functions, GPS data and others on the system in order to obtain more useful indexes.

6. References


