The Concept and Application Methods of Intelligent Content

Yong-Bae Yoon
Department of Computer Software
Sejong University, Seoul, Korea

Song-Hwa Chae
Information and Computer Engineering
Korea IT Industry Promotion Agency, Seoul, Korea

Won-Il Kim*
Department of Digital Contents
Sejong University, Seoul, Korea

ABSTRACT

Intelligent Content is defined as detailed information or fragment of content which contains a semantic data structure. This semantic structure makes possible to do various intelligent operations. There are wide range of content-oriented applications such as classification, retrieval, extraction, translation, presentation and question-answering. The concept of Intelligent Content is applied to various fields like MPEG and Semantic Web. In this paper, we discuss the several important researches of Intelligent Content and how to apply this conception to these fields.

Keywords: *semantic web, intelligent content, XML*

1. INTRODUCTION

Nowadays we are flooded with various contents such as game, music and movie. It is difficult to manage these contents if it is not organized and accessed. Most contents do not have whole of their information. However, sometimes users want to know about detail information from fragment of the content. Intelligent Content makes possible to know the simplest fragments of external content. In this paper, we discuss the definition of Intelligent Content and illustrate several examples of it. We define Intelligent Contents in Chapter 2 and discuss various kinds of content in Chapter 3. Finally, we conclude in Chapter 4.

2. DEFINITION

Intelligent Content is information content with explicit semantic description through machines and people can share meaning and value. Intelligent Content has semantically structured data. It widely used for the purpose of a content-oriented application such as classification, retrieval, extraction, translation, presentation and question-answering. The structured data provides those technologies with an exact semantic input [1]. The distinct characteristic of Intelligent Content is that knowledge is hidden behind every fragment of content. It is significance that Intelligent Content has knowledge of the whole information about content but it is also important how many information content linked.

The target of Intelligent Content is obtaining a good understanding to the relation between the type of media, the work flow, genre and a media style. The adaptation and transformation of content elements across various productions and platforms are easy to achieve.

It is necessary to analyze, study and develop the metadata that defines a parameter required for generation and operation of the semantics media of many type and an object for such a purpose, media semantics and ontology. Practical methods of context-based information retrieval will be researched that simplify the location and retrieval of images, sounds, videos, characters, movements or behaviors from very large datasets and media storage systems. We have to improve methods and tools for language processing and saying integration to supporting

* *Corresponding author. E-mail: wikim@sejong.ac.kr*
Manuscript received Jun. 27, 2006 ; accepted Sep. 08, 2006

This paper is supported by Seoul R&BD Program

generation of the contents of media that consists of many languages [3]. By the same lines, MPEG-7 is framework of content-based retrieval and browsing, among others. Incorporation of linguistic annotation into MPEG-7 is in the agenda, because linguistic descriptions already constitute a main part of existing metadata [7]. In the next section, we discuss several kind of Intelligent Content.

3. VARIOUS KINDS OF INTELLIGENT CONTENT

3.1 Web Content

Web Content has semantically structured data. In addition, it can share meaning and value between machine and people. Their characteristic of web content allows that web content is a kind of Intelligent Content.

The semantic web is information that is given well-defined meaning. It makes easy people to work in cooperation with computer. Content on the web has been shifting from exclusively human-oriented content to more and more data content. The semantic web is more efficient than current web. It use a data that is defined and linked by well defined methods that supports discover, automation, integration and reuse across various application.

Semantic Web provides intelligent services such as information brokers, search agents and information filters. Intelligent services on the Knowledgeable Web should get ahead of the currently available versions of these services, which are limited in their functionality, and only work as stand-alone services that do not interoperate. The Semantic Web, with machine process-able information contents, is only possible when further levels of interoperability are established. Standards must be defined not only for the syntactic form of documents but also for their semantic content. This semantic interoperability is facilitated by recent W3C standardization efforts, notably XML / XML Schema and RDF/RDF Schema. These efforts are summarized in Figure 1.

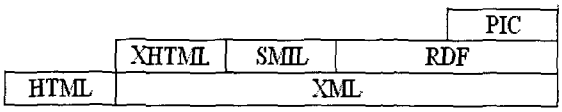

Fig.1. Language Layers on the Web

XML [10] is the basis of software development activities and widely known in the WWW community. XML is intended as a markup-language for arbitrary document structure, as opposed to HTML, which is a markup language for a specific kind of hypertext documents. An XML document consists of a nested set of open and close tags, where each tag can have a number of attribute-value pairs. The advantage of XML is that the vocabulary of the tags and their allowed combinations is not fixed. An example depicted in Figure 2 serializes a part of the ontology given above.

```
<class-def>
   <class name="plant"/>
   <subclass-of>
      <NOT><class name="animal"/></NOT>
   </subclass-of>
</class-def>
<class-def>
<class name="tree"/>
   <subclass-of>
      <class name="plant"/>
   </subclass-of>
</class-def>
<class-def>
    <class name="branch"/>
    <slot-constraint>
        <slot name="is-part-of"/>
        <has-value>
          <class name="tree"/>
        </has-value>
    </slot-constraint>
</class def>
```

Fig.2. Partial XML-Serialization of the Ontology

Fig.2 is the basic data-model of XML that is a labeled tree. Each tag corresponds to a labeled node in the data-model and each nested sub-tag is a child in the tree. It is important to point out that except OWL language it is only possible XML syntax express the ontology. We easily define many other XML versions of the same semantic information because of this character of XML.

Any XML document whose nested tags form a balanced tree is a well-formed XML document. Furthermore it is possible to enforce constraints on which tags should be used, and which nesting of these tags is allowed.

3.2 Multimedia Content

MPEG-7 and MPEG-21 are used by multimedia contents and there are several kinds of Intelligent Content. MPEG-7 describes the data and it manages in MPEG-21. MPEG-7 provides description of standard of core technology.

The MPEG-7 descriptions of content includes information describing the creation, production processes of the content (director, title, short feature movie) and the usage of the content (copyright pointers, usage history, broadcast schedule). Also, it has information of the storage features of the content (storage format, encoding), structural information on spatial, temporal components of the content (scene cuts, segmentation in regions, region motion tracking). Moreover it has much information about low level features in the content such as colors, textures, sound timbres and melody description. All these descriptions are coded in an efficient way for searching, filtering and so on [7].

Figure 3 illustrates a simple example of conceptual aspects description. The figure involves that Yuhki Kuramoto is playing the Piano. The event is characterized by a semantic time description: "7:30 PM on the 3th of October 2002", and a semantic place: "Seoul Arts Center". The description involves one event (to play), and three objects (piano, Yuhki Kuramoto,

and the abstract notion of musicians). The last two objects belong to the class of Agent.

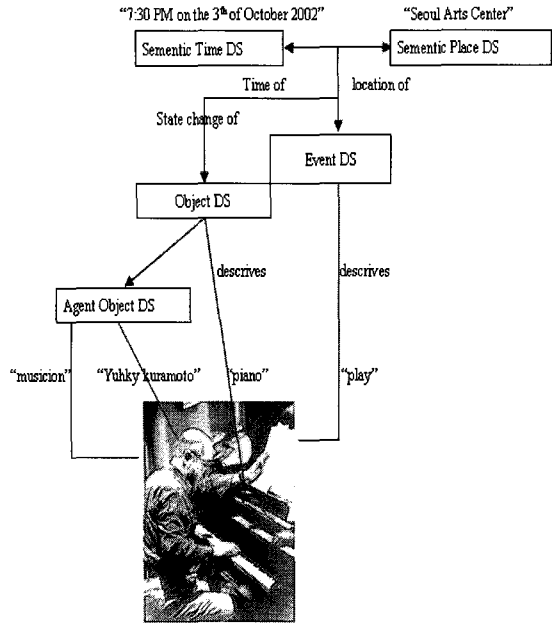

Fig.3. Example of conceptual aspects description.

MPEG-7 DDL (Description Definition Language) is a language for defining the structure and content of MPEG-7 documents. It represents the results of modeling audiovisual data. MPEG-7 has schema and defines a class of XML documents. MPEG-7 instances are XML documents that conform to particular MPEG-7 schema (expressed in the DDL) and describe audiovisual content.

We discuss simple example in Figure 3. It is possible to represent with MPEG-7 schema. Figure 4 shows how to describe Fig. 3 with MPEF-7 description.

```
<TextAnnotation>
    <FreeTextAnnotation xml:lang="en">
        Yuhki Kuramoto playing the Piano.
    </FreeTextAnnotation>
    <KeywordAnnotation>
        <Keyword>musicion</Keyword>
        <Keyword>Yuhki Kuramoto</Keyword>
        <Keyword>piano</Keyword>
        <Keyword>play</Keyword>
    </KeywordAnnotation>
    <StructuredAnnotation>
        <Who><Name>Yuhki Kuramoto</Name></Who>
        <WhatAction><Name>playpiano</Name></WhatAction>
        <Where><Name>Seoul Arts Center</Where></Name>
        <When><Name>7:30 PM on the 3th of October 2002</Name></When>
    </StructuredAnnotation>
</TextAnnotation>
```

Fig.4. A simple example of DDL with Fig.2.

MPEG-21 is the multimedia framework for management multimedia content. In order to build the substructure that supports multimedia content manipulation, content should be individually identified, described and managed._Transmission and transformation of content is provided to the terminal and network that are different in each other and the event generated this process. Also this process requires function of pursuit and report. Such report technologies include management of the individual data and the preference which try to be piling the user privacy and reliability transfer in the function.

MPEG-21 is the common substructure (multimedia framework) that can support all the situations concerning such content. Multimedia framework design was made to standardize at a notional dimension.

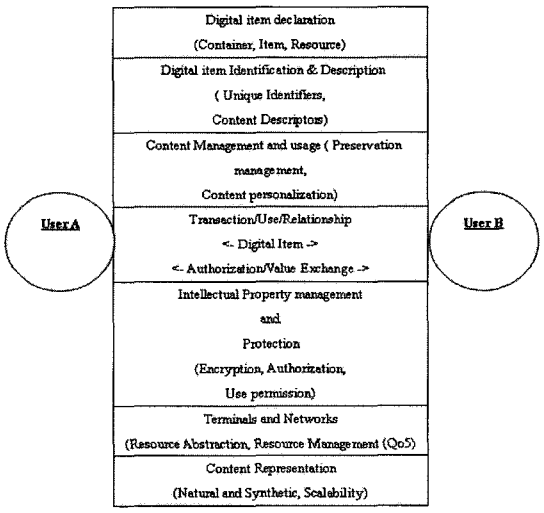

Fig.5. Composition element of a multimedia framework

Fig. 5 shows seven composition elements of a multimedia framework standard. MPEG-21 considers making requirement for making multimedia framework standard that users transmit, search, preserve and manage for proper multimedia contents. This framework gives a particular number and definition to every digital item. This number and definition is used for describing Digital item when it transfers and becomes a candidate for management previously. This number and definition are not discovered easily in everywhere. In addition, it enacts the standard about the protection of copyrights and can accommodate many kinds of writing right protection solutions in order to protect the right of the author of a digital item. This is the purpose of MPEG-21. It can be said to be the purpose of MPEG-21.

MPEG-21 describes their schema with DIDL (Digital Item Declaration Language). It provides an XML element set and XML Schema for representing a DID in XML. DIDL has a hierarchical structure, which makes it possible to give information on the structure of the digital item. For instance, one digital item can be represented by different instances in different formats (such as images in JPEG or TIFF), or a movie can be represented by separate clips that together make up the main feature.

3.3 Education Content

An Intelligent Educational Content (IEC) is a kind of Intelligent Content. It makes to understand what it is premature, content format. Figure 6 shows an example of IEC. It provides

the student's ongoing solution of a problem few additional coding [6]. There are very few domains where the knowledge behind a fragment of educational content that is automatically generated by the system from the problem statement. For example, following representation might be used in an IEC which teaches in C.

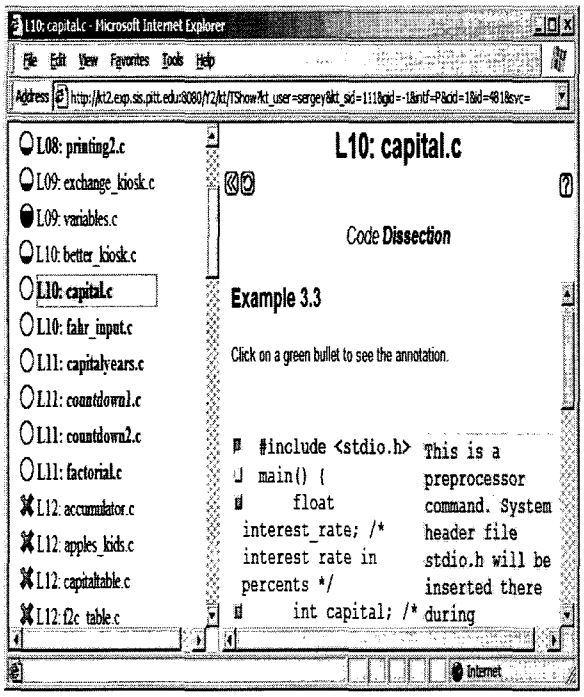

Fig.6. The Interface of NavEx

The NavEx content consists of three frames. At the left of frame contains shows a list of links among all examples and it makes possible to know a student's current course. The colored icons are annotated links. A green circle means that student has studied the examples. Also a red bullet means that the student has not mastered enough view the example. The link annotated with the red bullet is thus disabled. The order of links to examples is fixed, so that students can find them in the same place, no matter what the student's progress through the course has been. The shadow of L10 means that student study current page. The central frame displays the name of the current example. Underneath of the name, it has two links. One loads the source code of the example into the central frame. It shows where it will be copied, compiled and explored. The other link loads interactive example dissection. Dissection means that one takes the original source code and comments on it. These comments address the meaning and purpose of this line of code and help the student to understand the example better. Extended comments are shown to the left of the code and can be activated by clicking on the bullet next to the line of the code.

4. CONCLUSION

We discuss about Intelligent Content. Intelligent Content has semantically structured data and it used from such fields as computational linguistics, document processing, terminology, information science and multimedia content. Moreover it is very important that how combine the concept of Intelligent Content to each fields. Next research, we investigate method of combine Intelligent Content and fields.

REFERENCE

[1] Paul Buitelaar, Kôiti Hasida, "Semantic Annotation and Intelligent Content", Centre Universitaire Luxembourg, 2000.8

[2] Herb Simon, "A wealth of information creates a poverty of attention", 2004

[3] Herwig Zeiner, "Requirements for a common Multimedia Ontology Framework by JOANNEUM RESEARCH", 2006.1

[4] B.S. Manjunath, Philippe Salembier, Thomas Sikora, "Introduction to MPEG-7", JOHN WILEY & SONS LTD. Pub., 2002

[5] Ron Loborde, "Multimedia Management System for Intelligent Content Based Retrieval", University of Bristol, 2004

[6] Sergey Sosnovsky, Peter Brusilovsky, "Michael Yudelson, "Supporting Adaptive Hypermedia Authors with Automated Content Indexing", University of Pittsburgh, 2003

[7] José M. Martínez, "MPEG-7 Overview (version 10)", ISO/IEC JTC1/SC29/WG11N6828 Palma de Mallorca, 2004.8

[8] Katashi Nagao$_1$, Shingo soya$_2$, "Making the World Wide Web More Understandable and Usable with External Annotations", IBM Research$_1$, Tokyo Research Lab., Keio University$_2$, 2000

[9] Dr Mark H. Butler, "Is the semantic web hype?", Digital Media Systems Department HP Labs Bristol, 2006.3

[10] Stefan Decker1, Sergey Melnik 1, Ian Horrock 2, "The Semantic Web - on the respective Roles of XML and RDF", Stanford University1, University of Manchester2, 2001, Vol. 1, No. 4, 1991, pp. 249-272.

Yong-Bae Yoon

He studies Software engineering at Sejong University, Korea. His main research interests include Intelligence Content and Embedded Systems.

Song-Hwa Chae

She received a B.E. in Information and Computer Engineering from Ajou University, Korea in 1997 and a M.E. in Computer Engineering from Ajou University, Korea in 1999. She received Ph.D. in Information Communication from Ajou University, Korea in 2006. From 1999 to 2001, she was researcher of Korea Information Security Agency to study PKI system. Since 2006, she has been with Korea IT Industry Promotion Agency, Seoul, Korea. Her research interests include digital content, digital convergence, wireless communication system, and secure information systems.

Won-Il Kim

He received the B.E in Metal Engineering from Hanyang University, Seoul, Korea in 1982. He worked for Korean Air from 1981 to 1985 as System designer and programmer. He received the B.S., M.S in Computer Science from Southern Illinois University, U.S.A. in 1988, 1990 respectively. He received Ph.D. in Computer and Information Science from Syracuse University, U.S.A. in 2000. From 2000 to 2001, he worked for Bhasha INC, U.S.A. as technical and research staff. He was with Ajou University, Suwon, Korea from 2002 to 2003. Since 2003, he has been with Department of Digital Content, College of Electronics and Information, Sejong University, Seoul, Korea. His main research interests include Artificial Intelligence, Multimedia Contents and Computer Security.