2004－ＦＩ－74　(6)
2004－ＤＤ－43　(6)
2004／3／26

Digital Content Types based on Information Media Symmetry

Kunio Ohno
DoCoMo Systems, Inc.

ohno@docomo－sys.co.jp

Digital content or digital document is composed of information media elements with types. The types are related to the symmetry that exists between the history of information media that has been used for human communication and the history of information media that has been augmented with the expansion of computer application areas. The former reflects the inductive thinking of humans, while the latter corresponds to the deductive processing of computers. Type is the goal of the generic concepts that have been generated in the process of information abstraction in the induction process. It is also possible to think that the types in program languages, CORBA IDL, and XML Schema are their application to computer systems, network systems and document systems respectively.

1. Introduction

Information media are the media for facilitating human communication. The invention of letters and languages is a characteristic that distinguishes humans from other living creatures. Humans have steadily acquired these forms of information media on their way to the creation of today's advanced civilization. I first take a general overview of the information media that have been acquired by humans in the course of their history. The earliest humans had no proper language. It is conjectured that they delivered messages by shouting or making gestures. Humans later progressed to the sharing of information using pictures, as indicated by the prehistoric cave paintings discovered at Altamira in Spain and Lascaux in France. Around that time, humans also began to acquire a form of language. Since then, through a steady progression via hieroglyphs, pictographs, ideographs and phonograms, humans have evolved their current communication media.

Next, I discuss the progress of computer information media. Computers were first used for binary scale calculations, and then began to be used in areas such as technological calculations and business data processing. The available 'letters' were expanded from just alphanumeric letters at the beginning to kanji characters, and computers were introduced to non－Western cultural areas, including Japan. Thanks to the progress of semiconductor technology, computer systems then became more compact and displayed higher performance. As a result, the use of workstations and PCs was no longer restricted to specialists, but spread to the general public. Moves were then initiated to make computers easier to use rather than just to increase their performance, and they became capable of handling figures, pictures and even multimedia.

The first forms of human information media used the visual and audio senses and then expanded the process to the use of pictures, hieroglyphs, pictographs, ideographs and phonograms. The computer expanded its information media from the use of binary to alphabets, kanji characters, figures, pictures and multimedia. It occurred to me one day that these two evolutionary paths seemed to correspond, but were moving in opposite directions. That was why I began the research for this paper. Having specifically considered this point further, I have

found that the media history of humans and computers do indeed correspond in a remarkable way. The specific details and the reasons behind such a development follow.

2. The progress of human information media

Table 1 is a summary of the information media that have been used by humans right from their earliest days up to the present. Here, I'd like to discuss the idea that the progress of these media revolved around the basic characteristic of humans to move towards the generalization and abstraction of individual pieces of information.

Table 1. Communication media used by humans

PERIOD	MEDIA
100,000 BC～	Shouts, gestures
30,000 BC～	Cave paintings
5000 BC～	Hieroglyphs, pictographs
3000 BC～	Ideographs
1500 BC～	Phonograms
1000 AD～	Algebra
1800 AD～	Modern logic

2.1 Communication through shouts and gestures

The earliest humans had no language as such. It is thought that their basic communication media were shouts and gestures. Some time later, to fulfill the need to cooperate in various ways, such as when making tools and using them effectively, humans became to seek more efficient forms of communication. The fact that they were making tools means that there must have been a purpose assigned to each tool, so the process required to achieve a particular purpose had already been recognized by humans at that time. It is possible that some simple form of language was used to specify fellow humans, animals, and places, and also to give directions for particular actions [1].

2.2 Cave paintings

Cave paintings from around 30,000 BC that still exist today are an invaluable part of the human artistic heritage [2]. Particularly well known are the paintings of animals that have been found in South Africa (29,000 years old) and the cave paintings of Altamira in Spain and Lascaux in France (between 13,000 and 25,000 years old) [3]. It is thought that these paintings were used for ritual purposes. It is certain that the humans who painted them were involved in some kind of communication designed to improve the management of their particular group.

2.3 Hieroglyphs and pictographs

Hieroglyphs and pictographs came into use around 10,000 BC. The Sumerian cuneiform characters dating back to around 6000 BC appear to have been developed from pictorial symbols used for bookkeeping purposes [4]. It seems that people began by drawing simple pictures on clay boards to record crop and livestock data. To do that, they needed either to sharpen the tip of a stick or make it wedge−shaped. It appears that cuneiform characters were eventually born when people found that pressing the wedge−shaped tip of a stick onto a clay board to make marks was far more efficient than drawing pictures.

The Egyptian hieroglyphs are the best known and most popular in the world, and many books on them are still being published today [4]. There were initially (c. 4000 BC) around 700 different pictorial symbols that were used in a wide range of fields, including agriculture, medicine and education, as well as law. More hieroglyphs were created to meet particular needs, and the total number had grown to around 5,000 by the end of Cleopatra's reign in 27 AD.

The kanji characters of the Yellow River civilization also began as pictorial symbols [4]. But their number was increased according to systematic rules. They are still used today in several countries, including Japan.

2.4 Ideographs

Cuneiform characters and Egyptian hieroglyphs became extinct along with other hieroglyphs and pictographs, and only kanji 'characters' survived. They were further developed into ideographs. Because they originated from specific items, hieroglyphs were suitable for referring to objects, but ideographs had to be made capable of more general and more abstract expressions. In order to express a new concept, new characters had to be created or existing characters had to be combined. As a result, new ideographs were continually being added, and the total number of characters became enormous. Idioms were also defined by creating compounds. However, it has never been easy to learn all those letters and idioms. It is sometimes pointed out that one merit of kanji characters is that you can read and make yourself understood through writing even if you don't speak the language. But, as a trade−off, kanji have many intrinsic problems, the greatest being the time required to memorize the huge number of characters. That makes it extremely difficult to increase a nation's literacy rate.

2.5 Phonograms

Phonograms are often regarded as being originated from alphabets, but that is not necessarily true. It is known that phonograms using hieroglyphs were already in use before alphabets. The Japanese kana syllabaries are known as a typical form of phonogram. As seen in the case of the 'Manyo' type, Japanese kana were initially expressed in the orthography of kanji, simply borrowing phonemes from them. Later, kana phonograms were formed by simplifying kanji characters (hiragana) or extracting portions of kanji characters (katakana) to correspond to vocal sounds.

It seems that alphabets were also formed from hieroglyphs through a process fairly close to that of kana. A Phoenician form of writing that spread to Greece is regarded as the origin of the European alphabets. In the case of the English alphabet, it became possible to read and write simply by learning the basic 26 letters. That was really a revolution, considering that there are 600 kinds of cuneiform letters, between 700 and 5,000 kinds of hieroglyphs, and several tens of thousands of kanji characters. The adoption of an alphabet contributed greatly to the diffusion of writing and an increase in the literacy rate within races as well as nations [4].

If we define world history in a holistic fashion, it is considered that the fact Western civilization was able to become the mainstream of the world was strongly related to the adoption of alphabets for its writing systems. Similarly, Japanese kana have contributed to the increase of the literacy rate in Japan, and it is believed that their invention greatly helped Japan to achieve a status comparable to the advanced nations of the West.

2.6 Algebra

The 'number concept'−sum, remainder, product, (quotient), etc.−was initially developed in Egypt. The Egyptians discovered that there are 365 days a year (or 366 days once every 4 years) largely by observing floods on the River Nile, and they developed surveying technology out of the need to make crop forecasts. That is how the numerical formula concept of the four rules of arithmetic was formed [10]. Although calculations using specific numbers was widely known among various ancient civilizations, it was not until the 9th century AD that a systematic concept of variables (algebra) was born, in which numbers are handled abstractly. The Arab mathematician Muhammed ibn Musa al−Khowarizmi is regarded as the founder of algebra. The name 'algebra' seems to be derived from the title of his book, Hisab al−jabr w'al−muqabalah, published around 830 AD [5].

2.7 Modern logic

The Greek philosopher Aristotle systematized 'classical logic', known for its syllogisms. Strictly speaking, it cannot be termed a 'system' because it is based on the meaning−oriented concept of natural languages. Modern logic, unlike classical logic, is logic that has been system-

atized in an algebraic method structured with both the concept of 'sets' and the logic of symbolism.

Mathematics was re−systematized by modern logic. Gottlob Frege tried to restructure algebra in accordance with the theory of sets [6]. However, his approach ran aground due to the paradox discovered by Bertrand Russell−the essential contradiction of a proposition of which the object is a set itself. The logic for overcoming that paradox is the 'Type Theory', which is regarded as a model for the 'data type' in computer science [7]. Russell later systematized algebra and co−authored Principia Mathematica with A.N.Whitehead.

3. Progress in information media handled by computers

The information media handled by humans have made rapid progress thanks to the advances in communication technology since the Industrial Revolution, such as the invention of the telegraph and the telephone, and broadcast technology during the 20th century, including radio, television, and recording, as well as playback technologies, such as photographs, gramophones and movies. Based on such technologies, the media handled by computers have also expanded gradually. This evolution is summarized in Table 2.

Table 2. Media handled by computers

PERIOD	MEDIA
1940s	Binary system (Machine language)
1950s	Numerical calculation (FORTRAN)
1960s	Alphanumeric characters (COBOL)
1970s	Kanji processing (Japanese system)
1980s	GUI (icons, mouse)
1990s	Figures, pictures
2000s	Video, sound

3.1 Machine language system with binary system

ENIAC (Electronic Numerical Integration And Calculation), completed in 1946, is generally regarded as the first computer, although there are differing opinions [8]. It is certainly true that it was one of the first machines that actually worked as a computer. ENIAC's program used a patchboard wiring connection, but it was not the computer that A.M.Turing had advocated for the realization of both operations and data in one and the same language system. The first practical and genuine 'Turing Machine' was the EDSAC (Electronic Delay Storage Automatic Calculator), completed in 1949. It was the first computer to use accumulated programs and realize the basic functions of today's computers, including a set of basic commands and subroutines.

3.2 Numbers & numerical formula calculation

In the 1950s, as semiconductors came into use for the basic computer elements, we entered the age of large, centralized computers. The first field of application was technological calculations. Fields that had been unsolvable by any conventional form of manual calculation−such as the numerical analysis of differential equations and the finite element method−were opened up to contribute to the progress of science and technology. Major examples of such application fields were the various systems of the U.S. Department of Defense, the design field of the manufacturing industry, and the academic world.

3.3 Alphanumeric character & business data processing

In the 1960s, computer applications using TSS (Time Sharing System) turned towards business data processing. Prior to that, alphanumeric codes (including ASCII) had been established and were put to use in fields where the computer had already been introduced. That spread to the field of business data processing.

The text editor for program development was diffused, and was then applied to the fields of general word processing and office data processing. Later, COBOL (Common Business−Oriented Language) emerged, and it was established as the standard program language for business data processing at government offices and in industrial fields.

3.4 Kanji processing

In the 1970s, a Japanese word processing system was introduced. First, the kanji code (JIS−6225, 6226) was established and then the kanji system was developed by major computer manufacturers (IBM, Fujitsu, Hitachi, NEC and Denden Kosha, now NTT). Furthermore, wire dot and laser beam printers were developed for outputting kanji. Another important event in the same decade was the introduction of microprocessors, which led us towards the present age in which the cost performance ratio of personal computers and word processors is constantly improving. Riding on a wave of such technological innovation, the first Japanese word processors were introduced in the latter half of the 1970s. After that, Japanese computer and household electric appliance makers began producing and marketing Japanese word processors one after another. It was not long before hand−written office documents were a thing of the past and word−processed documents were the norm.

3.5 Graphical User Interface (GUI)

The Palo Alto Research Center (PARC) of Xerox Corporation, developed the GUI means of communication with a computer, in which icons and windows are manipulated by using a mouse [9]. It was introduced to workstations and personal computers in the 1980s. The origin of this technology was the Alto machine developed at Xerox PARC. The multi−window system was developed, using program languages like Smalltalk and Interlisp−D, and a GUI adopting icons and a mouse was introduced. Another result of that project was the Ethernet, that links computers at high speed. Object−oriented technology, based on Smalltalk and high−speed LAN, became the starting point for the distributed object technology that supports the network culture today.

3.6 The age of figures and pictures

In the 1990s, along with the trend towards lower memory prices and the spread of object technology, systems for figures (2D and 3D) and pictures (CG) became available for use in cooperation with the conventional documentation system. Adobe began marketing Illustrator for figure−data processing and Photoshop for image−data processing. Kodak began offering technology such as PhotoCD to transfer conventional silver halide film photos to CDs. These developments have lead to the birth of new occupation fields, such as 'computer artist.'

Digital contents with figures and pictures are spreading to every corner of our lives today, thanks to lower prices, higher performance, the diffusion of the Internet, the use of maps on car navigation systems and the Web, and the diffusion of digital cameras.

3.7 The age of video, sound and animation

In the new millennium, under the 'Broadband' slogan, humans have entered into an age of full−scale multimedia, where video, sound, and animation are fully utilized. Meanwhile, broadcasting, communications, and information processing have become very closely connected and are now forming a single market. We can say that digital TV and the Web will merge, with XML (eXtensible Markup Language) as the key technology.

Considering the inevitably higher speeds of the Web in future, broadcasting is expected to merge with the Internet. Broadcasting will shift to cable, and it is very likely that radio will shift to the side of mobile communication applications, targeting mainly individuals and mobile communication media. As bandwidths broaden, images and animation will also become fully utilized in the mobile communications world.

4. Comparison of information media progress between humans and computers

4.1 Communication versus computers

Table 3 shows the way in which Tables 1 and 2 correspond with each other. The information previously shown in Table 2 is here presented in reverse order. Table 3 indicates the interesting idea that the history of communication in human society and the history of computer processing correspond beautifully—in opposite directions. Is this just a coincidence? Or is there any logical consequence behind it? I believe there is a logical consequence, as I will explain below.

Table 3. Media comparison: human communication versus computers

HUMAN	COMPUTER
Shouts, gestures	Video, sound
Cave paintings	Figures, pictures
Hieroglyphs, pictographs	GUI
Ideographs	Kanji processing
Phonograms	Alphanumeric characters
Algebra	Numbers
Modern logic	Binary theory

4.2 The history of abstraction in human communication

We can say that the progress of human communication media has been a history of generalization, establishing models, creating symbols, and making abstractions. Starting from shouts and gestures, humans gradually acquired words and language. They invented letters so that they could communicate a greater number of meanings and contents using a smaller number of marks. Furthermore, in order to study universality and objective truth, humans systematized the way of thinking in disciplines such as mathematics and logic. This continuous process can be regarded as the history of human intellectual development.

4.3 The history of instantiation in computers

Compared to the history of human communication, the progress of computers can be described as a history of instantiation, visualization and objectization. These processes evolved with the progress of computer science and cost reductions for memories and communication. For a human being, it is far easier—and, indeed, preferable—to grasp meaning and contents directly without having to depend on logical thinking. Humans who prefer logic and mathematics are rare, and most people naturally prefer watching a movie of a story in a short time to reading the complete story in book form, which takes more time. Those who brought this idea to the computer world were the LRG (Learning Research Group) people, including Alan Kay of PARC. It has produced many businesses that have contributed to the promotion of multimedia.

5. Discussion

5.1 Induction and deduction

I hope it is now clear that the correspondence indicated in Table 3 is not a coincidence and that, in fact, a logical consequence exists there. In other words, the logical thinking of abstraction—oriented humans and the computer's deductive approach to data do correspond [20]. From a logical point of view, these two processes are referred to respectively as 'induction' and 'deduction.'

Extracting the same or similar natures by an inductive method in various phenomena is the process regarded as the concept formation process. Putting a name to such concepts and making them the basis for classification was the very model of the birth of ideographs and new words.

Religion and philosophy have been two approaches to the systematization of such induc-

tive ways of thinking to help define the world viewed by humans. Plato's theory of 'ideas' and Aristotle's 'metaphysics' are classic examples. It can also be said that modern analytical philosophy, including logical mathematics, has tried to systematize such ideologies and philosophies from a stricter mathematical viewpoint.

5.2 The constructive and analytical approaches

It was in 1906 that Bertrand Russell presented his 'Type Theory' to overcome his 'Paradox.' The basic concept of the 'type' was described in Annex B of his book, The Principles of Mathematics (1903). Russell clearly presented his idea of 'type (set)' and its 'member (instance)' regarding numbers, thus explaining an idea close to the 'data type' of today. Though The Principles of Mathematics has been overshadowed by the Principia Mathematica (co−authored by Whitehead, 1910?13), it gives us a much better idea of what he was really challenging in constructing this world according to logic.
Russell explains human intelligence from the aspect of the 'understanding of numbers'. He presented it as follows: humans can define this world using either the 'constructive approach' or the 'analytical approach'. In the former, the process consists of 'number→ infinity→ space→ substance→ movement', while in the latter, the process is 'symbolic logic→ description theory→ type theory.'

5.3 The possibility of two directions

As described above, The Principles of Mathematics tried to systematize the phenomena of the world using symbolic logic (set theory). But Russell met with the famous paradox on the way, and he could not develop his ideas logically any further. The Principia Mathematica could be regarded as Russell's challenge once again to describe this world after overcoming the paradox. However, he was challenging the logical definition of the world, and I think his basic ideas are described better in The Principles of Mathematics than in Principia Mathematica.

He presented a way of thinking to develop the idea of 'type'−developing things based on numbers−by dividing it into two directions−constructive and analytical. The two directions here seem to correspond to two ideas: the idea of developing data into multimedia and simulation; and the idea of constructing a world that expresses meanings using metadata and metamodel−oriented system construction.

5.4 The correspondence between type/class and media

Figure 1 shows the correspondence between the relationship shown in Table 3 and the relationship of media and the conventionally used 'type/class':

It should be clear how modern logic, algebra, phonograms and ideographs correspond to the 'types'−including boolean, int, float, char and wchar−which have been established along with the progress of the computer. The fact that hieroglyphs and pictographs are here made to correspond to the 'icon−class' in the object−oriented GUI system might be a slightly distorted interpretation. As for the figures and pictures of cave paintings, shouts, gestures and multimedia, they are defined here to be the application−dependent and implementation−dependent types and class of object. However, there is a possibility that they will be standardized as the type to be recognized commonly in various areas in the future. Thus the 'type' is the human heritage that was formed as humans made abstractions−via the induction process−of typical things available to be used in common through communication. In such circumstances, it is to be expected that reusable frameworks of various areas will continue to be constructed steadily from now on.

6. Conclusions

I have discussed the symmetry that exists between the history of the information media that have been used for human communication and the history of the information media that have

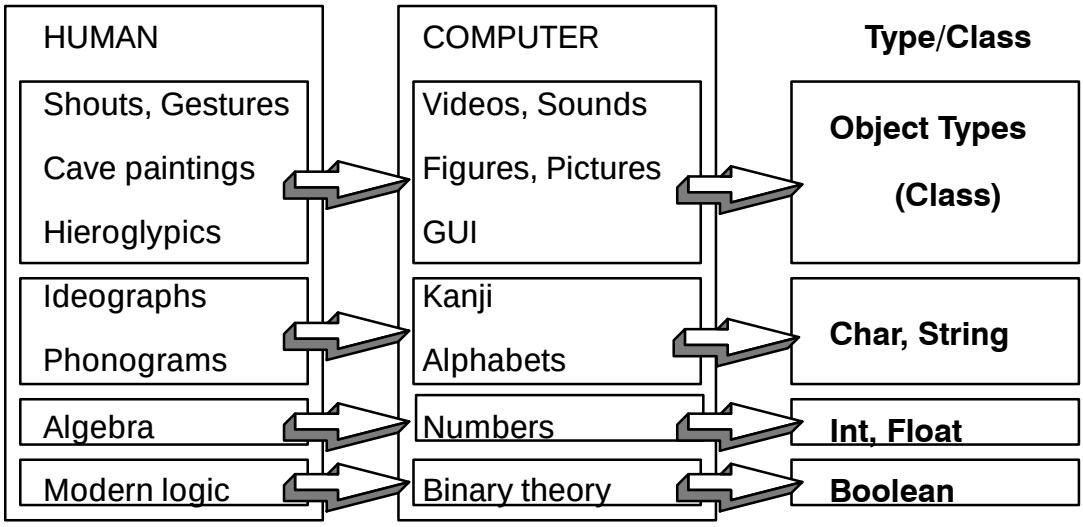

Figure 1: The Relationship between Various Media and Type/class

been augmented by the expansion of computer application areas. The former reflects the inductive thinking of humans, while the latter corresponds to the deductive processing of computers. 'Type' is the goal of the generic concepts that have been generated in the process of information abstraction in the induction process—that is to say, the process of metamodeling. It is also possible to think that the 'types' in program languages, CORBA IDL (Interface Definition Language), and XML Schema are their application to computer systems, network systems and document systems respectively.

References

[1] C. Stringer, R. McKie, African Exodus—The Origins of Modern Humanity, Brockman, New York, 1996.

[2] R. E. Leakey, The Making of Mankind, Dutton , 1981.

[3] G. Jean, Langage de Signes, L'Ecriture et Son Double, Gallimard, 1989.

[4] G. Jean, L'Ecriture, Memoire des Hommes, Gallimard, 1987.

[5] D. Guedj, L'Empire des Nombres, Gallimard, 1996.

[6] B. Russell, Introduction to Mathematical Philosophy, Allen & Unwin, 1919.

[7] C. J. Cleaveland, An Introduction to Data Types, Addison—Wesley, 1986.

[8] C. R. Mollenhoff, ATANASOFF : Forgotten Father of the Computer, Iowa State University Press, 1988.

[9] L. Tesler, "The Smalltalk Environment", Byte, Vol. 6, No. 8, 1981.

[10] B. Russell, The Principles of Mathematics (2nd Ed.), Routledge, 1937.