Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 797-805
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

Exploring and Practicing the Integration of ZCC Technology into Digital Archives

Xin Cui[1], Yuhang Cui[2*]
[1]School of Economics and Management, Huangshan University, Huangshan City, Anhui Province, China
[2]Business School, Hehai University, Nanjing, JiangSu, China
[*]Corresponding Author.

Abstract:

Based on the analysis of the current situation of digital archive system construction and ZCC technology, this paper analyzes the feasibility of the integration of digital archives and ZCC technology, and designs the path of the integration of the two. The project of digital archive renovation of enterprise A is selected as a sample, to explore the effect of integration of ZCC technology into digital archive system, and solve the technical problems of large image files in digital archive systems, such as high-fidelity compression/decompression, encryption of image files, fast transmission, retrieval and utilization. In the era of the knowledge economy, the archive, as an important knowledge resource, is a strategic resource of enterprises and an important tool to enhance their core competitiveness. While the development of digital archives is not achieved overnight. With the rapid development of technology, it is bound to bring challenges to the use of the original digital archive system. The fundamental driving force of enterprise archival work is to provide quality service. How to meet the realistic working needs of enterprises and serve their development more effectively requires us to explore the transformation of enterprise archiving systems. As new technologies such as block chain, big data and artificial intelligence continue to emerge, incorporating new technologies has become an important topic in the area of archival modernization. Based on the implementation and research of enterprise A's digital archive renovation project, this paper explores and addresses the problem of the integration of ZCC technology into digital archives.

Keywords*: ZCC technology, Digital archives, Large image files, Technology and system integration.*

I. INTRODUCTION

1.1 Current Research of Digital Archives

In the 1970s, with the first batch of electronic documents received by the US National Archives, the construction of digital archives began to be explored. In the 1980s, the US National Archives and Records Administration (NARA) established an electronic document center, which began to collect, categorize and store electronic documents and provide access to them. In the 1990s, the University of Virginia Library established the Jefferson Digital Archives, after which the digital archives began to develop rapidly. The construction of digital archives in China was explored on a pilot basis from 2000 to 2005, which focused

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 797-805
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

on the development of archival machine-readable catalogue formats and electronic document archiving and management specifications (national standards), implementation program planning, software development, hardware configuration and digital processing. From 2006 to 2010 it was in the development stage, and construction of electronic documents center and digital archives was started. The construction was fully launched from 2011 to 2013, with the main focus on the use of archival resources. In 2014, the China Archives Bureau issued *the Testing Method for Digital Archive System*, based on the concept of "business-oriented and use emphasized" and the pursuit of the combination of "archives management and business activities". The service performance has become increasingly prominent. By 2018, with the application of technologies such as blockchain, big data and artificial intelligence to digital archives, intelligent archives has become an upsurge of research in the archival industry.

1.2 Current Research of Image Compression Technology

The current research on image compression technology is focused on international standards for image compression and image coding algorithms, the goal of which is to minimize storage space without losing image pixels, so as to maximize network transmission speed. Image compression technology originated in the 1940s, and since Claude Shannon proposed Information Theory, image coding technology has developed rapidly, with the emergence of Shannon-Fano coding, Huffman coding, dictionary coding and predictive coding[1], etc. In the 1980s, the International Organization for Standardization, the International Electronically Commission and the International Telecommunication Union-formed Joint Photographic Experts Group (JPEG) to develop a general image compression standard [2]. In the 1990s, with the establishment of Wavelet Theory, Fractal Theory and Visual Simulation Theory, image coding technology took a quantum leap forward. In 1997, JPEG established the static sequence image compression standard JPEG2000 [3]. In the 21st century, the image coding method ushered in a new upsurge based on wavelet theory, and researchers have proposed many innovative coding methods, such as the Curvelet Transform, the Minimum Path Wavelet Transform and the Hadamard Transform. In addition, the fractal coding [4] method has become a new hot topic in wavelet theory research. In 2003, the ITU and ISO/IEC jointly introduced the H.264/AVC video coding standard [5]. In 2010, the ITU-T and ISO/IEC started to develop the LE High Efficiency Video Coding Standard (HEVC). After the HEVC coding framework was proposed, it has received the attention and participation of many research institutes. At present, a series of achievements have been made in the research of image coding technology.

1.3 ZCC Technology Principle

The ZCC (zero-loss compression and retrieval of color document format) is a compression technology based on Wavelet Theory, and is the result of many years of research by a Chinese technology company. In image compression processing, the wavelet transform image classification and fractal color image compression coding algorithms are adopted. The image to be compressed is divided into several sub-images according to the frequency, and they are divided into several levels according to the analysis of the correlation of the three color components of the color image, the hierarchical information redundancy of the fractal quad tree coding and the number of coding contained in each image, and different

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 797-805
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

compression algorithms are adopted, respectively. Three independent color components of the image are combined into one in some way to search for matching blocks, so it is necessary to reduce the three-color component matching blocks stored and searched to one, and compress the quad tree hierarchical information. As the low-frequency sub-image with the largest coding amount as the first level, the lossless Differential Pulse Code Modulation (DPCM) is adopted; the high-frequency sub-image containing a small amount of coding is called the second stage, and then the embedded zero-tree coding method is adopted; for higher-frequency sub-images with little coding, they contain little coding and can no longer be encoded.

Using different combinations, a variety of coding methods with similar image compression ratios and decoding quality are obtained. The experimental results show that it is better than the SFC method and the standard JPEG method. It has the advantages of a high compression ratio, fast transmission and download speed, and can finish coding at any time without affecting the image quality, and can be reversed at any time. Because ZCC technology uses image separation technology, the image color blocks are analyzed before image compression, and the images are separated according to different color structures. In this way, an image with high magnification and high compression is obtained.

II. MATERIALS AND METHODS

2.1 Keyword and Paper Search

In terms of literature research, a subject search was conducted with “Digital Archive,” “Image compression,” “Digital Archive + image compression” as search terms as of March 20, 2022. The results of the search using the Web of Science database were 8755 entries, spanning the period 1982 to 2022, with the number of entries increasing from 11 in 1997 to 603 in 2021; the distribution of topics was in information science, library science, computer science, archival information systems and archival information resources. Similarly, 60582 results were found for “image compression”, spanning the period 1963 to 2022, with the number of results increasing from 100 in 1990 to 3661 in 2021, with a clear trend of growth, focusing on wavelet transform, image coding, image compression algorithms and JPEG. On the other hand, an advanced search for “Digital Archive* Image compression” found zero entries. From the analysis of the retrieval results, we can see that there has been widespread interest in the fields of digital archives and image compression, but there is a lack of discussion on the integration of the two, and this study is expected to strengthen this weak link.

2.2 Analysis of Integration Demand

At present, the difficulty in the construction of digital archives lies in the storage and transmission of high-quality color images. If images are in high-quality color mode, the file is large, which makes it difficult to transmit, browse, download and use on the network. Therefore, these valuable information resources cannot be effectively used. The deep integration of ZCC technology with digital archives will help solve these problems and facilitate the implementation and advancement of archival information.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 797-805
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

2.3 Demand for High Fidelity Image Compression and Decompression

At present, the main storage formats for high-fidelity images are JPG, PDF, TIFF and ZCC.

The JPG format is a way to save color images. It is not suitable for multi-page electronic document management. JPG cannot be embedded in OCR recognition and is not suitable for secondary development and reuse. It can be applied to high fidelity color electronic documents, but it takes up much space for storage.

With the internationally popular JPEG2000 format, compression, reconstruction and restoration of smaller images can be achieved without loss of image quality, without optimization. But JPEG2000 encoding takes a long time. And large image processing is a lossy compression, in which the same image will be blurred after several forwards.

The PDF file format has been widely adopted internationally as a document browsing format. The "Technical specifications for the digitization of paper archives of the National Archives Administration of China" require that the storage format be TIFT, JPG format, and the images used for network query are CEB and PDF [6]. A PDF is divided into image PDF and text PDF. The image PDF embeds the scanned image into the PDF, and the color image compression still adopts the JPG algorithm. Although PDF is a multi-page single-document structure, multi-page JPG files embedded in a single document PDF will make the PDF large.

TIFF can be divided into two types: lossless and lossy. Although lossless TIF can save color image information well, it takes up a lot of storage space and cannot meet the requirements of file network utilization services. The lossy TIF format has a high compression rate, but it can only compress electronic files into black and white, which means that the originality of the electronic files is lost and it cannot be used in the archival business.

In recent years, ZCC technology has also been applied to the construction of digital archives, which has made outstanding achievements in enhancing and improving the processing and utilization of large images in digital archives, and enabling a better user experience. Assuming that the input document size is a standard A4 format with a resolution of 30pdi, the hardware configuration used in the test was an ordinary computer with a CPU of 2.0HZ and a memory of 512MB, and the results of the experiment are shown in Table I.

TABLE I. ZCC, TIF, JPEG and PDF Comparison of advantages and disadvantages

FORMAT / TECHNOLOGY	**TIFF**	**ZCC**	**JPEG**	**PDF**
TYPICAL FILE SIZE BW	50-100KB	5-30KB	WITHOUT	50-100KB
COMPRESSION TIME BW	≤1S	1.5S	WITHOUT	≤1S
COMPRESSION RATE BW	1X	2X	WITHOUT	1X

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 797-805
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

TYPICAL FILE SIZE COLOR	2500KB	30-100KB	500-2000KB	WITHOUT
COMPRESSION TIME MULTI	WITHOUT	3-6S	2S	2S
COMPRESSION RATE COLOR	1X	300-1000X	12-50X	12-50X
IMAGE QUALITY	HIGH	HIGH	LOW	LOW
PLUG-IN SIZE	WITHOUT	1MB	0KB	40MB

By comparing the experimental data, it can be concluded that the ZCC format has greater advantages in terms of color high fidelity, image quality and color file size. Its compression algorithm makes the image quality of the ZCC format more stable and better than other image formats, which is the main reason it is easier and more convenient to manage large image files.

2.4 The Need for Image Encryption, Fast Browsing, and Full-Text Retrieval of Text Information in Image Files

Most of the construction of archival digitization uses scanners to digitize images. While because images, especially large drawing images occupy a large storage space, the speed of transmission, retrieval, browsing and downloading is slow in the network environment. It cannot meet the urgent need for timely, fast, accurate and convenient use of archival resources.

In terms of practical project, when dealing with large images, we take the following steps. The first step is to convert the electronic file into a raster image file in memory through virtual printing technology. This converts the image into a ZCC format. The second step is to convert the image to ZCC format while conducting OCR recognition of the text information in the image, which resides in the ZCC format file. Then "by searching engine technology, invalid text is filtered, search rules are established, and relevant XML file description rules are established at the same time as the search, so that the full text information in the image can be searched" [7]; the third step is to achieve fast browsing. "An A4 format using 24-bit true color 300 DPI, scanned output into TIF format with file size of 24718K, PDF 2653K, JPG 2430K" [8], while the ZCC format is only 23K. ZCC format files are transmitted on the network due to paging data streams, and they can appear in only 3 seconds in the narrowband 56K network environment, which improves the speed of image browsing.

2.5 Design of the Integration

According to the demand analysis, the integration of ZCC format technology into the existing digital archives can improve its value in use. Regarding how to integrate, we need to analyze the target digital archive system, and design the integration path from four aspects: user layer, application layer, database service layer and business logic layer.

The first method is to add algorithms and achieve integration through the various module interface programs provided by the original system. This method requires full support from the original system developers. In practice, as the programs provided by the original system to the user units are encapsulated,

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 797-805
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

direct modification to add algorithms is unfeasible. In addition, the original developers are unwilling to share technology, therefore it is difficult to achieve system integration.

The second method is to create a new data chain and add an image database, only requiring the original system user module to provide an external interface program, which can be solved by introducing an Agent. While the adoption of more agents will involve the problem of multi-agent cooperation and a large amount of development.

The third method is to add a new algorithm to where the interface can be provided by the original system. In modules where the interface is not available, the Black Box Method is adopted, with the agent linked to the periphery to run the new function. This integration method can get the interface and add a new algorithm. If it cannot get the interface, it can be designed by building a data link and introducing Agent.

Based on the above considerations, this paper adopts the third approach, proposes an integration model (as in Figure 1) and constructs an integration path.

The so-called Agent is a computer system that can receive and transmit information, perceive the environment and react in a timely manner. It is able to perceive the environment and "has characteristics such as cooperation, initiative, autonomy and self-adaptability. It can connect and work with other Agent, respond instantly to the environment" [9], and be initiative-taking and interactive when accepting an instruction or service. Agent technology provides an effective way to solve new distributed application problems.

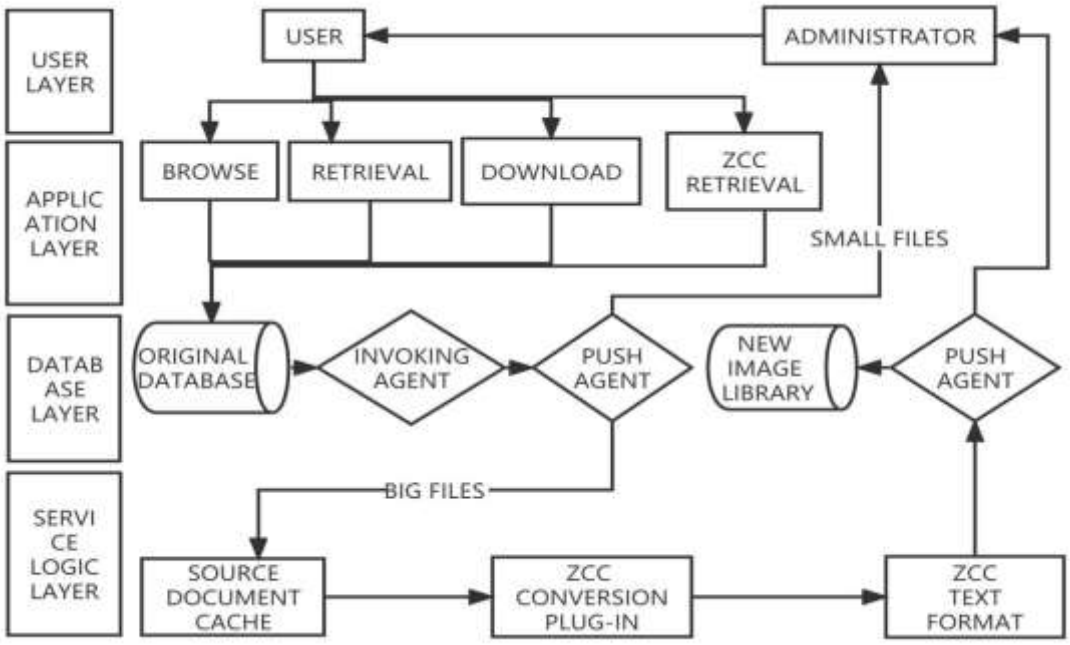

Fig 1: ZCC integration path

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 797-805
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

2.6 Discussion on the Application Layer Integration

At the application layer program interface, the ZCC image search module is added. When users need to retrieve relevant image documents in the digital archive system, they can click on the ZCC image search box to directly retrieve image database data, and through its path they can realize multi-level in-depth search and directly query keywords in drawings, images and multi-page text reports to meet the needs of users for fast and accurate query and utilization of archival materials. The original full-text search and browse queries can be used simultaneously, but only small files can be retrieved from the original database.

2.7 Discussion on the Integration of the Business Logic Layer

ZCC technology can be deployed in the following three ways: ①Install the ZCC conversion software on the server, and deploy the conversion mode on the server, where the client uploads the source file to the server, and the server unifies the compression and converts the image into ZCC format; ②Install the conversion software on the client, and the client automatically uploads it to the server after conversion; ③Provide the ZCC technology controls. After the user transfers the file to the controls, the controls convert it, and then output it to the user. All three of the above options can be chosen to achieve business process configuration. In this paper, the third approach is used in the design of the integration. When the user needs to download a file, the Processing Agent is triggered. By judging the file attributes, the small files are sent to the user through the Agent. The large files are sent to the file cache. Then, the ZCC conversion plug-in is activated to convert them to ZCC format and send them to the Push Agent, where the new image database is updated and stored, and at the same time, the user is sent the files to finish downloading them.

III. CONCLUSION

The construction of digital archives needs to keep pace with the times, and the fundamental purpose of its optimization and upgrading is to realize the organic combination of new technologies and the deployment of the original system, which is a basic project to realize the rapid co-construction, sharing and efficient utilization of information resources. Through the summary of the practical project, it is considered that the main effects are reflected in the following aspects.

3.1 Significant Reduction in Hardware Investment

The object of this study is the large archives of enterprise A. If the original system is used, the archives will need to add hardware such as servers, storage, data backup and network interfaces every year. After the integration of ZCC technology, only a few servers need to be added for hardware deployment, storage and data backup hardware do not need to be added, and the network bandwidth can be maintained as it is, which greatly reduces the maintenance cost of the archives.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 797-805
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

3.2 Improve the Efficiency of Utilization Services

If the push agent to the ZCC plug-in is used, this will realize the secondary digital processing of the large image file, instead of the administrator's manual processing and improves the processing efficiency of large image files. At the same time, the experimental results show that: ZCC format occupies smaller storage space. Compared with other formats, TIFF, PDF, and JPEG are ZCC format multiples are: 1074 times, 115 times and 105 times respectively. Using ZCC technology not only saves time for users to browse and download, but also improves user satisfaction. After the implementation of the original system transformation of ZCC technology, the 12-month data statistics showed that the year-on-year growth was 58% and 64% respectively, and the use efficiency was significantly improved.

3.3 Problems Existing in Integration

After introducing ZCC technology into the original system, the archives organization still needs to solve some problems in practice, mainly including: (1) the precise management and security of physical files; (2) the integration with the digital file system will inevitably make the technology adaption difficult for the original users; (2) The integration of new technologies will inevitably affect management models. For archives managers, they will face many challenges in management, innovation of archives service methods, and service efficiency.

3.4 Implications

This study has practical value in terms of innovative paths for integrating new technologies into the original data archive system. It begins with a needs analysis, followed by the selection of an integration method and the design of an integration path, and finally the evaluation of the integration effect. This paper is based on a single case study, which limits the generalization of the findings. To make the technical path of this paper more solid, more research methods, such as multiple case comparison studies, can be used to verify the findings of this paper and strengthen the explanatory power of the process of realizing the innovation path proposed in this paper. This will help open the "black box" of how to integrate new technologies with the original system.

ACKNOWLEDGEMENTS

This research was supported by National Natural Science Foundation of Anhui Province, China (Grant No. KJ2015A165).

REFERENCES

[1] Li ZN, Drew MS, Liu J. Fundamentals of multimedia. Upper Saddle River (NJ): Pearson Prentice Hall; 2004.
[2] Cui SX. System Construction of Chinese Archives Protection Technical Standards. Archives Science Bulletin. 2007(1):62-7.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 797-805
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

[3] Boliek M. JPEG2000 part I final draft international standard. (ISO/IEC FDIS15444-1), ISO/IEC JTC1/SC29/WG1 N1855. 2000.
[4] Yuen CH, Lui OY, Wong KW. Hybrid fractal image coding with quad tree-based progressive structure. Journal of Visual Communication and Image Representation. 2013 Nov 1; 24(8):1328-41.
[5] Wiegand T, Sullivan GJ, Bjontegaard G, Luthra A. Overview of the H. 264/AVC video coding standard. IEEE Transactions on circuits and systems for video technology. 2003 Aug 4; 13(7):560-76.
[6] ZCC (2022) ZCC "Zero" Loss, High-compression Color Document Technology. Available at: http://www.zccsoft.com/list.php?typeid=17.
[7] Zhang XK. Several issues about the digitization of paper archives in university archives. Lantai World, 2013(S3):103.
[8] Gul S, Bano S. Smart libraries: an emerging and innovative technological habitat of 21st century. The Electronic Library. 2019 Oct 7.
[9] Wooldridge M, Jennings NR. Intelligent agents: Theory and practice. The knowledge engineering review. 2018 Jun; 10(2):115-52.