

Optimization of electronic document management systems by means of encoding and visualization of stored data in the integrated development environment

*Mikhail V. Poskonin, Andrey O. Kalinin, Igor V. Kovalev, and Mikhail V. Saramud**

Reshetnev Siberian State University of Science and Technology, 660037 Krasnoyarsky Rabochy Av. 31, Russia

Abstract. The article considers the issue of storage, cataloging and transfer of documentation within the electronic document management system (EDMS). Modern document management solutions can be considered as the intersection of traditional and digital technologies, in particular, the storage of software code on paper. The very storage of program code in paper version implies a large archive for data storage, which includes modifications when versions of the program code are updated. Graphical ways of storing large amounts of data have been known for a long time, but they usually use a long procedure for restoring the original information and exclude recovery from paper or are significantly limited in the amount of stored data. As one of the elements of the composite electronic document management system it is proposed to use the method of graphical encoding of information (QR-code generation technology), which will increase the resistance to unauthorized modification of the program code, simplify the procedure for storing, identifying and verifying information. As an elementary block of information, it is intended to use a separate software module responsible for one of the subprograms in the RTOS environment. The proposed algorithm for storing and transferring information implies its compression by software, encryption and transmission. The article compares various compression algorithms and their efficiency. The presence of a cataloguer in each graphic element allows for quick search, comparison and verification of blocks.

Introduction

Modern workflow solutions should be considered at the intersection of electronic and paper technologies [1]. The heterogeneous structure of these systems allows for a different

* Corresponding author: msaramud@gmail.com

relationship between electronic and paper documents, depending on the sphere of application, user requirements and other parameters of the workflow [2].

The degree of influence of the components of the structure is determined by the level of technology development and the stability of the document workflow traditions. Document management systems can function successfully provided that the contradictions between the views of the customer and the developer about system are overcome, and the balance of the use of paper and electronic documents will be found [3]. These circumstances determine the composite approach to the implementation of document management systems, where both purely electronic and purely paper documents coexist, as well as many composite solutions.

Today, basically, users and creators of EDMS operate intuitively, creating systems of varying degrees of efficiency and sustainability that do not provide continuity.

The basis of the proposed approach is: principles of automation of organizational management systems [4], principles of object-oriented design [5] and principles of project management [6].

One of the main ideas is formulated in the principle of automation of workflow [4]. The essence of this principle is that EDMS should be designed not only as a medium, but as an instrument for active interaction with documents. In the object-oriented approach, this is commonly called encapsulation. [5]. Successful implementations based on correct technical solutions, but not taking into account existing practice of using the facility, will not be successful.

As one of the elements of composite EDMS, it is proposed to apply the method of graphical encoding of information (QR-code generation technology), which will increase the resistance to faulty changes in the code, simplify the procedure of storing, identifying and verifying information. As an elementary block of information, it is intended to use a separate software module responsible for one of the subprograms in the RTOS environment.

As the basis of the graphic part, a monochrome twenty-module QR code with a minimum level of error correction is proposed [7, 8].

For the organization of EDMS program code, a client-server interaction structure is proposed. The server is responsible for storing information, and determining the access rights of clients to the content of the server. The client side performs authorization requests and provides the ability to read, edit and generate QR codes.

According to the authorization and developer instructions, the program code is assigned the following attributes:

- the author of program code;
- the co-authors of the developer;
- the name of the program code;
- version of the program code;
- a timestamp.

Upon completion of the development of the program code for uploading it to the server, the source code is processed:

- removal of extra spaces and blank lines, deletion of comments and optimization of program code (these manipulations are optional, i.e. implemented at the request of the developer);

- using irreversible encryption algorithms, such as MD5 or SHA-1, we get the HASH-sum of the program code that provides data integrity control and is added to the end of the program code (fig. 1);

- in the process of text compression, regular expressions are reduced to their minimized form, compilation of dictionaries and other actions to reduce the size of the encoded information (fig. 2);

- using data encryption and compression algorithms such as the Huffman algorithm, Shannon-Fano algorithm, arithmetic coding, LZ methods (named after Abraham Lempel and Jakob Ziva), the initial data is compressed and encoded [9, 10];
- the modified data is converted into a graphic image by applying the QR-coding algorithm (fig. 3);
- based on the HASH-sum a unique name, which serve as a cataloging element for quick searching and working with files, is assigned to a graphic file. The name itself can be present on a paper copy of the element in the form of a bar code;
- the finished image with the specified attributes is downloaded to the server, where a corresponding entry is made to the database (fig. 4, 5).

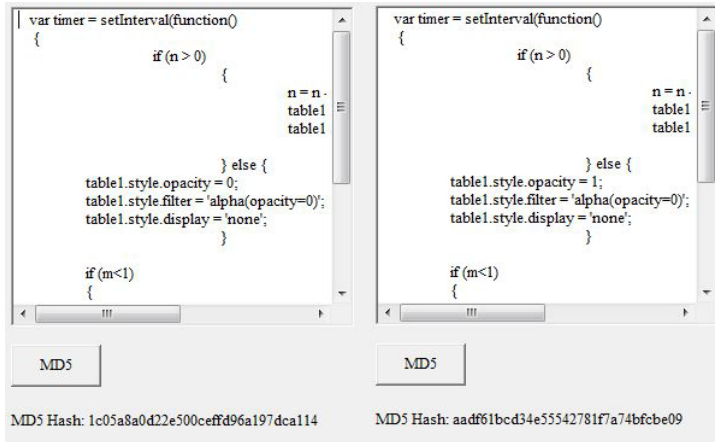


Fig. 1. Hash-sum when one symbol changes in the program code.

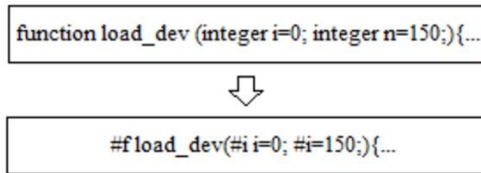


Fig. 2. Example of regular expressions.

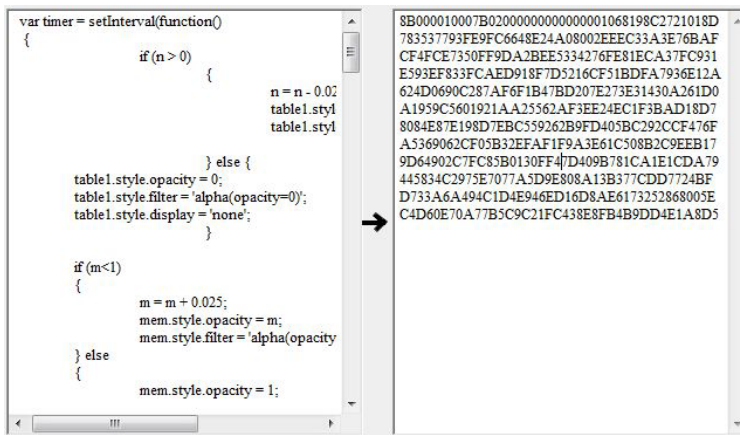


Fig. 3. Compression and Data Encryption.



Fig. 4. Graphical display of the element name of the EDMS database.

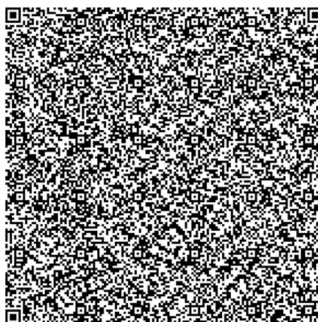


Fig. 5. Graphic display of the compressed program code consisting of 8429 symbols.

The decryption process is carried out in the reverse order.

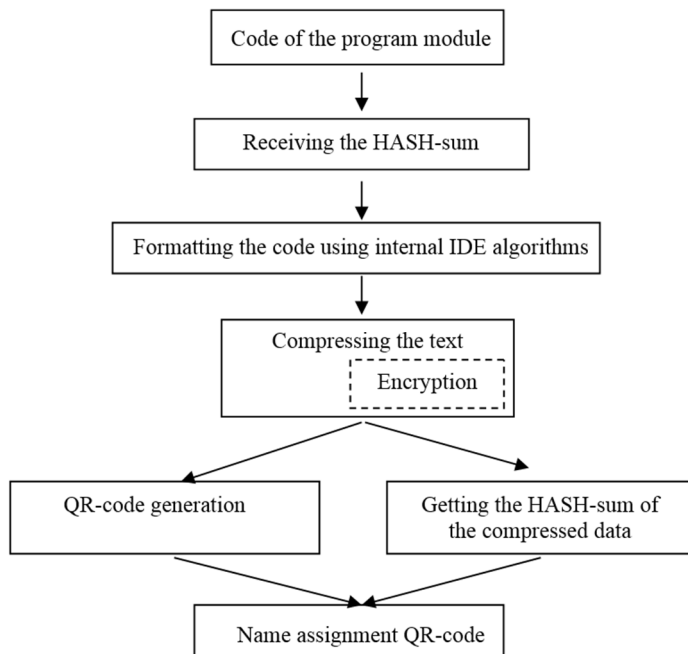


Fig. 6. Document life cycle in EDMS.

Several classical text compression algorithms were selected as a data compression algorithm. The efficiency of algorithms was analyzed, the results obtained are given in Table 1.

Table 1. Results of comparing data compression algorithms.

The ratio of the input data volume to the output, expressed as a percentage		The number of symbols supplied to the input of a compression algorithm		
		2 000	10 000	50 000
The algorithm of data compression	Huffman	127%	85%	32%
	Shannon-Fano	93%	87%	86%
	LZMA	65%	54%	52%

The efficiency of various compression algorithms is determined by the different amount of input data. In this regard, we can conclude that the need to select a compression algorithm based on information about the amount of input data. Since we consider a program module as a stored unit, the size of which, as a rule, exceeds 30 thousand symbols, the Huffman algorithm is the most optimal.

Encryption can be the next step, which is optional in this system. In the case of documenting modules that are not of particular importance, compression algorithms and graphical encoding of information appear as a sufficient measure of protection. In the case of storage of modules of special importance or commercial secrecy, encryption may be additionally applied.

Since we have a limit on the amount of information stored in each QR-code, situations may arise when the block size exceeds this amount. In this case, the block is divided into a sequence of several elements, in each of which is added an indication of what part it is by the number and from how many. This information is also placed in an explicit form on paper form for the convenience of document processing.

When processing one of these documents, the system informs the user that the Nth document from M has been processed, and suggests downloading the remaining fragments. Only after this, the whole document is decrypted, checksum comparison is performed, etc. Thus, we actually remove the limit on the size of the stored program block, if its size increases, it will only increase the number of documents necessary for its storage.

Conclusion

The result of this work is the functional structure of a composite electronic document management system that is responsible for processing, storing and monitoring versions of software code in the life cycle of the software under development. The life cycle of the document in the proposed EDMS is described. Comparative analysis and selection of compression algorithms used in the proposed system are made. Data storage and verification, taking into account the implementation of the integrated EDMS structure, allows to restore the original information when the electronic copies of the archive are corrupted, and also simplifies its processing and storage.

This work was supported by Ministry of Education and Science of Russian Federation within limits of state contract № 2.2867.2017/4.6

References

1. Leikums, Toms. International Journal of Advances in Engineering & Technology, 5.1 194, (2012)

2. L. C. Smith, and M. A. Wong, eds. *Reference and Information Services: An Introduction* (Libraries Unlimited, Santa Barbara, 2016)
3. M. Sögtrop, U.S. Patent No. 8,892,532. 18 Nov. (2014)
4. K. Dalkir, *Knowledge management in theory and practice* (Elsevier, Routledge, 2013)
5. M. Warasart, and K. Pramote, 4th International Conference on Computer Engineering and Technology. International Proceedings of Computer Science and Information Technology, 94 (2012)
6. B. G. Grabkowitz, and J. C. Nicholas " U.S. Patent No. 8,736,878. 27 May (2014)
7. A. Walsh, *The Electronic Library*, 29.4, 429 (2011)
8. P. Kieseberg, et al. Proceedings of the 8th International Conference on Advances in Mobile Computing and Multimedia. ACM (2010)
9. R. Ashford, QR codes and academic libraries: Reaching mobile users (Faculty Publication – George Fox University Libraries, 2010).
10. S. R. Kodituwakku, and U. S. Amarasinghe. *Indian journal of computer science and engineering*, 1.4, 416 (2010)