Section 2

COMPARISON OF ARCHIVE FORMATS
APPLICABILITY FOR PRACTICAL USE

Andrii Litvinov

National Aerospace University «Kharkiv Aviation Institute»

Scientific adviser: Artem Perepelitsyn

Relevance. In the era of large-scale data management, archiving formats have become essential tools for reducing storage costs, optimizing file transfer efficiency, and enhancing data organization. As digital data continues to grow, selecting an optimal archive format has practical implications for both individuals and organizations [1]. Choosing the appropriate format can save storage space, reduce transfer times, and improve overall data handling efficiency, particularly in scenarios where bandwidth and storage resources are limited. This study addresses these needs by analyzing commonly used archive formats to determine their effectiveness in practical, everyday applications.

The purpose of this study is to conduct a comparative analysis of popular archive formats, specifically ZIP, RAR, 7z, and TAR.GZ, to assess their practical utility in terms of compression efficiency, data handling features, and suitability for a range of use cases. By evaluating factors such as compression ratio, processing speed, and recovery options, this study aims to guide users in selecting the best format for their specific archiving requirements.

Research results. The analysis of ZIP format compatibility shows support by various systems and acceptable compression efficiency. While standard ZIP has a 4GB file size limit, the ZIP64 extension allows larger archives. Despite its user-friendly design, ZIP lacks advanced recovery features, which limits its reliability for sensitive data applications [2]. RAR format shows a high compression ratio and multi-volume support, making it suitable for large datasets [3]. It also includes data recovery features for added resilience against corruption. However, the RAR format has only extraction source code available for the use that can be a disadvantage in open-source environments [4]. The 7Z format, supported by the open-source 7-Zip software, achieves strong compression through the LZMA algorithm. It is ideal for archiving large files due to its high efficiency and open-source flexibility. However, 7Z offers limited recovery options and may lack native support on some systems, which can be a drawback in standardized environments. Lastly, TAR, often paired with GZIP to form TAR.GZ, is common in UNIX and Linux systems. While TAR does not compress files on its own, TAR.GZ provides effective compression while

preserving file attributes. However, TAR.GZ lacks native encryption and data recovery features, limiting its suitability for applications requiring high data protection.

Conclusions. The performed investigation allowed us to identify the strengths and limitations of popular archive formats. ZIP and TAR.GZ are ideal for general use and compatibility, while 7z and RAR provide higher compression, suited for storage-limited environments. For open-source and high-efficiency needs, 7z is recommended, while RAR is preferable for situations requiring data integrity protection and data recovery. Selecting the right format depends on balancing compression, features, and compatibility with specific use cases.

List of references

1.   Prabavathy B., Ramya P., Babu C. Optimized private cloud storage for heterogeneous files in an university scenario. 2013. 3rd International Conference on Recent Trends in Information Technology (ICRTIT), Page. 323–328. DOI: https://doi.org/10.1109/icrtit.2013.6844224.

2.   Wei Y., Zheng N., Xu M. An Automatic Carving Method for RAR File Based on Content and Structure. 2010. 2nd International Conference on Information Technology and Computer Science (ITCS 2010), p. 68–72. URL: https://doi.org/10.1109/itcs.2010.23.

3.   Radescu R., Barar A. P. The Performances of the Fixed Constraints Transform Applied in Text Compression Experimental Results and Comparisons. 2018. 10th International Conference on Electronics, Computers and   Artificial   Intelligence   (ECAI),   Page   1–6.   DOI: https://doi.org/10.1109/ecai.2018.8678935.

4.   Guo Y., Li G., Zhang G. Design of the Scientific and Technological Information Management Software. 2023. 5th International Conference on Electronic Engineering and Informatics (EEI), Page 235–239. URL: https://doi.org/10.1109/eei59236.2023.10212809.

Information about the authors

Andrii Litvinov, student at the Computer Systems, Networks and Cybersecurity Department, National Aerospace University «Kharkiv Aviation Institute», a.a.litvinov@student.csn.khai.edu

Artem Perepelitsyn, Associate professor at the Computer Systems, Networks and Cybersecurity Department, National Aerospace University «Kharkiv Aviation Institute», PhD, a.perepelitsyn@csn.khai.edu