271. THE PROBLEM OF LINK ROT IN DIGITAL INFORMATION

Pishchyk S.I.
Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Maksimchuk R.T. – Senior Lecturer

The issue of link rot as a challenge for digital information preservation is examined in this paper. Factors causing link rot are analyzed. Modern approaches to combating link rot are evaluated. Recommendations for academic work are proposed.

The problem of broken links is becoming increasingly important in today's digital world. Link rot happens when web links no longer point to their original content. Every day, countless valuable online resources disappear when websites are closed or content is moved. This problem affects the quality of academic research and digital culture. According to studies, a significant proportion of web links become inaccessible within a few years after publication.

Link rot and content drift often occur together. Content drift is a situation where a resource remains accessible but its content has changed. Together, these two problems are called reference rot.

The scale of the problem has been documented across multiple domains. Research conducted at the California Institute of Technology found that more than 5 % of links to scientific data are inaccessible, with data disappearing at a rate of approximately 2.5 % annually. Resources identified by Digital Object Identifiers (DOIs) are lost three times less frequently than those referenced by standard URLs [1]. A study from Pew Research Center found that a quarter of webpages that existed between 2013 and 2023 are no longer accessible.

The causes of link rot include domain expiration, server shutdowns, and content reorganization [2]. Addressing these causes is essential for developing effective long-term preservation strategies. These factors are particularly problematic in academic contexts, where long-term access to sources is essential for verification and further research.

There are several ways to deal with link rot. Web archiving, for example, the Internet Archive's Wayback Machine, maintains archived copies of web pages. This service has preserved over 800 billion web pages since its founding, providing access to content that would otherwise be lost. This makes the Wayback Machine one of the most valuable resources for preserving digital heritage. Persistent identifier systems, such as DOIs, provide permanent references that remain valid regardless of changes to a resource's physical location. DOIs support the FAIR (Findability, Accessibility, Interoperability, Reusability) principles, ensuring unique identification and reliable access to objects. Distributed systems, such as IPFSs (InterPlanetary File System), store content across distributed networks, ensuring that data remains available even if individual servers fail [3]. This approach fundamentally changes how digital content can be preserved for future generations. These technologies offer new possibilities for long-term data storage that do not depend on a single organization.

Each approach has distinct advantages and limitations. Web archiving provides content preservation but depends on the continued operation of archival institutions. Persistent identifiers require infrastructure maintenance and registration fees, which may limit their use to formally published works. Distributed systems face challenges related to scalability and adoption rates. Choosing the right approach depends on the specific needs of a project or publication. Despite these limitations, combining different approaches can provide more reliable protection against link rot in academic publishing.

Based on the analysis of existing literature, several recommendations for students and researchers can be formulated: use DOIs instead of URLs when possible, archive local copies of important resources, check links before publication, and prefer resources from institutional repositories over personal websites. These practices are easy to implement and can be adopted by any student or researcher without additional costs. By following these simple steps, researchers can contribute to a more sustainable digital information ecosystem.

In conclusion, link rot represents a serious challenge to the long-term accessibility of digital information, particularly in academic contexts. Existing solutions require consistent application by authors and institutions. Simple preventive measures can substantially reduce the impact of the problem on academic work.

References:

1. Federer, L.M. Measuring data rot: An analysis of the continued availability of shared data from a Single University / L.M. Federer [et al.] // PLoS ONE. – 2024. – Vol. 19, № 6.

2. Teixeira da Silva, J.A. Archiving website-based references in academic papers: Problems caused by reference rot, potential solutions and limitations / J.A. Teixeira da Silva, M. Nazarovets // Learned Publishing. – 2023. – Vol. 36, № 4. – P. 1-12.

3. Reilly, L.J. Reilly EternaMark (REM) Protocol – Dual-Layer Digital Permanence Using DOI Archiving and Blockchain Timestamping // IETF Internet-Draft. – 2025. – URL: https://datatracker.ietf.org/ (date of access: 01.03.2026).