

Towards a Unified Demoscene Dataset: Consolidating Fragmented Archives

Abstract

The demoscene is a global subculture focused on creating real-time multimedia art demonstrations that showcase programming, artistic, and musical skills. Recognized as UNESCO Intangible Cultural Heritage in several European countries, its metadata is scattered and often duplicated across multiple online archives such as Demozoo, Pouet, CSDB, and others. This paper reflects on the process of building a unified dataset that consolidates metadata from these demoscene archives. By transforming the data into a standardized format and employing deduplication techniques based on external links and similarity, I have created a comprehensive, freely accessible dataset of demoscene productions. This dataset not only preserves the cultural and historical artifacts of the demoscene but also facilitates research in digital art, media studies, and computer science. It aims to support scholars, historians, and enthusiasts in exploring the rich tapestry of the demoscene movement.