

Link Management Tool with Internet Archive Integration

Graduate



Thomas Zahner



Benny Joe Villiger

Introduction: In today's digital age, the World Wide Web has become the indispensable platform for accessing and sharing information. At the heart of the Web are hyperlinks, also known as URLs or simply links. They are the threads that connect web pages and allow users to traverse the landscape of online content.

However, online content tends to be short-lived, and changes often go unnoticed. The phenomena of disappearing content and unannounced changes are known as link rot and content drift. Rotten or dead links are URLs pointing to resources that show obvious signs of malfunction. Content drift can result in linked information that is misleading or that differs significantly from the original intent of the linker. They can undermine the accuracy and credibility of websites, degrade the user experience and damage the reputation of the website or organisation. Checking and fixing links manually on a regular basis is time-consuming and error-prone.

Approach: We develop a command-line tool called Link Management Tool (LMT) providing a semi-automatic approach to detecting and fixing rotten links and drifted content. Additionally, we extend lychee, a free and open-source link-checker. We integrate the Wayback Machine, a digital Web archive, into LMT and lychee to obtain access to archived snapshots of websites taken in the past.

Result: LMT was released to the public on GitLab under the MIT licence. Its metadata snapshot functionality enables the detection of potential content drift on webpages. The modular design of LMT allows integrating it into scripts, providing the flexibility to schedule the link checking process at strategic intervals such as before each publication or release. Scheduling the link checking supports the continuous monitoring of links.

In addition, lychee was extended with a new feature that offers recommendations for broken URLs by utilising snapshots from the Wayback Machine.

Early user feedback and testing confirmed the reliability of LMT to identify and resolve broken links.

Advisor

Prof. Dr. Olaf Zimmermann

Co-Examiner

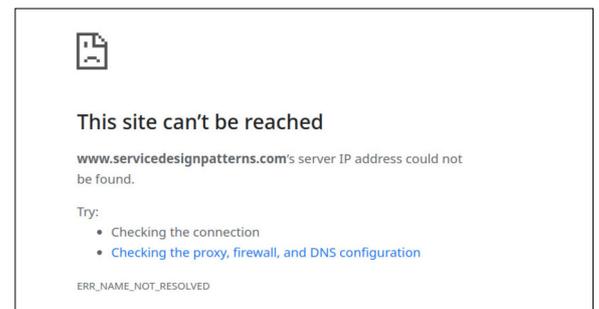
Dr. Hans-Peter Hoidn

Subject Area

Software, Internet Technologies and Applications

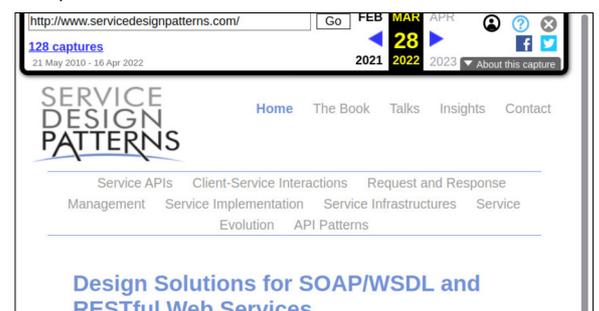
A broken link which leads to an unreachable site

Own presentation



The same broken link retrieved in the Wayback Machine

Own presentation



LMT in action: Find and fix broken links in a completely automated manner

Own presentation

