



Bern University
of Applied Sciences



URL-Archiver

Project report

Course of study

Author

Advisor

Co-advisor

Project 1

Nicolin Dora, Abidin Vejseli & Kilian Wampfler

Dr. Simon Kramer

Frank Helbling

Version 0.1 of September 30, 2023

- Technik und Informatik
- Informatik

Abstract

One-paragraph summary of the entire study – typically no more than 250 words in length (and in many cases it is well shorter than that), the Abstract provides an overview of the study.

Contents

Abstract	ii
List of Tables	v
List of Figures	vi
Listings	vii
1. Introduction	1
1.1. Initial Situation	1
1.2. Project Goal	1
1.3. Priorities	1
2. Specification	2
2.1. System Delimitation	2
2.1.1. System Environment (statics)	2
2.1.2. Process Environment (dynamics)	2
2.2. Requirements	2
2.2.1. Functional requirements (added value)	2
2.2.2. Boundary and Pre-Conditions	2
2.3. Usability	2
2.3.1. Personas	2
2.3.2. Storyboard	2
2.3.3. UX-Prototyping	2
3. Implementation	3
3.1. Architecture (e.g., back-/frontend)	3
3.2. Processes	3
4. Deployment/Integration	4
4.1. Installation (Sysadmin) Manual & Script	4
4.2. User Manual	4
5. Conclusion	5
5.1. Discussion	5
5.1.1. Example from BFH Template - Delete	5
5.2. Bottom Line	5

5.3. Future Work	5
Bibliography	6
A. Original Project Description	7

List of Tables

List of Figures

Listings

1. Introduction

1.1. Initial Situation

1.2. Project Goal

1.3. Priorities

2. Specification

2.1. System Delimitation

2.1.1. System Environment (statics)

2.1.2. Process Environment (dynamics)

2.2. Requirements

2.2.1. Functional requirements (added value)

2.2.2. Boundary and Pre-Conditions

2.3. Usability

2.3.1. Personas

2.3.2. Storyboard

2.3.3. UX-Prototyping

3. Implementation

3.1. Architecture (e.g., back-/frontend)

3.2. Processes

4. Deployment/Integration

4.1. Installation (Sysadmin) Manual & Script

4.2. User Manual

5. Conclusion

5.1. Discussion

5.1.1. Example from BFH Template - Delete

What is the significance of your results? – the final major section of text in the paper. The Discussion commonly features a summary of the results that were obtained in the study, describes how those results address the topic under investigation and/or the issues that the research was designed to address, and may expand upon the implications of those findings. Limitations and directions for future research are also commonly addressed.

5.2. Bottom Line

5.3. Future Work

Bibliography

A. Original Project Description

URL-Archiver

Description

The goal of this project is to deliver a FLOSS-licensed, platform-independent Java-program (called "URL-Archiver") that

- (1) takes as input (the path of) a directory or any Unicode-text- (e.g.: .BIB, .TEX; .HTML; etc.) or .PDF-file (<https://www.baeldung.com/java-curl>);
- (2) scans it for any URLs (<https://stackoverflow.com/questions/4026614/extract-text-from-pdf-files> , <https://librepdf.github.io/OpenPDF> , <https://pdfbox.apache.org> ; see also https://en.wikipedia.org/wiki/List_of_PDF_software);
- (3) extracts all URLs (regular expression ;-) from the text;
- (4) optionally spring-loads all URLs in a Web-browser;
- (5) posts all URLs to <https://archive.ph> ;
- (6) gets the resulting archived URLs;
- (7) outputs a CSV-file of the resulting key-value (URL, archived URL) pairs; and
- (8) optionally inserts the archived URLs into a .BIB-file.

The program code should be minimal, modular, and self-explaining.

The project report should be concise (maximally informative, minimally long).

It must contain this project description as a quotation.

Technologies

Java, LaTeX

Advisor

Dr. Simon Kramer

Declaration of Authorship

I hereby declare that I have written this thesis independently and have not used any sources or aids other than those acknowledged.

All statements taken from other writings, either literally or in essence, have been marked as such.

I hereby agree that the present work may be reviewed in electronic form using appropriate software.

September 30, 2023



A. Vejseli



N. Dora



K. Wampfler