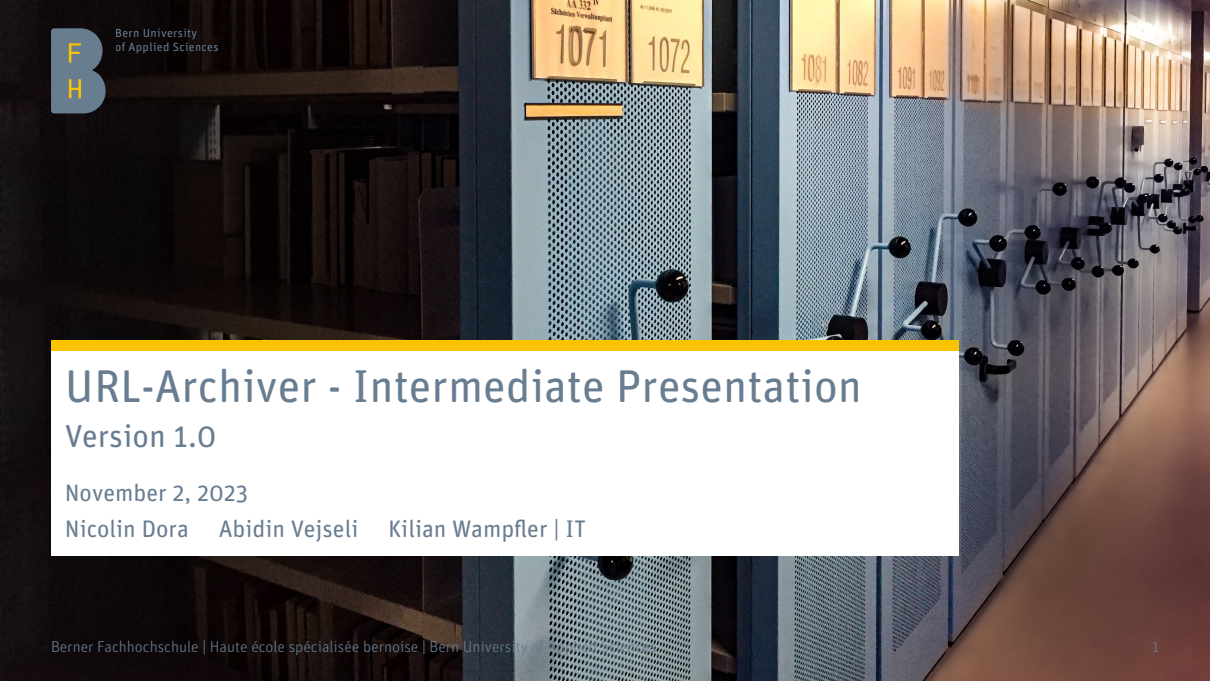


Bern University
of Applied Sciences

URL-Archiver - Intermediate Presentation

Version 1.0

November 2, 2023

Nicolin Dora Abidin Vejseli Kilian Wampfler | IT

Table of Content

We will speak about:

Problem Statement

Objectives

Requirements

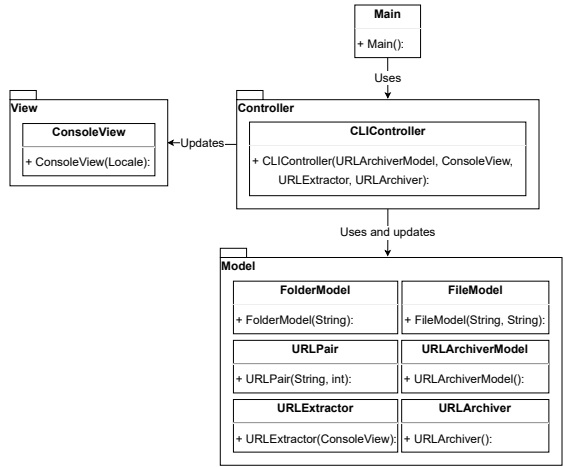
General conditions

Solving The Problem

Architecture

Employing the MVC Pattern

- Modular design for easy extension.
- Separate data, interface, and control flow.
- Facilitates the potential addition of a GUI.



Architecture

SOLID Principles

- Scalable and robust design.
- Ensures maintainability and clear code structure.
- Commitment to best programming practices.

Architecture

Object-Oriented Approach

- Promotes code reusability.
- Logical class hierarchies.
- Clear relationships between modules.

Architecture

Multilanguage Support with ResourceBundle

- Ready for global adaptability.
- Future-proofing architectural choice.
- Potential to accommodate multiple languages.

Data model

Structured Data Logic

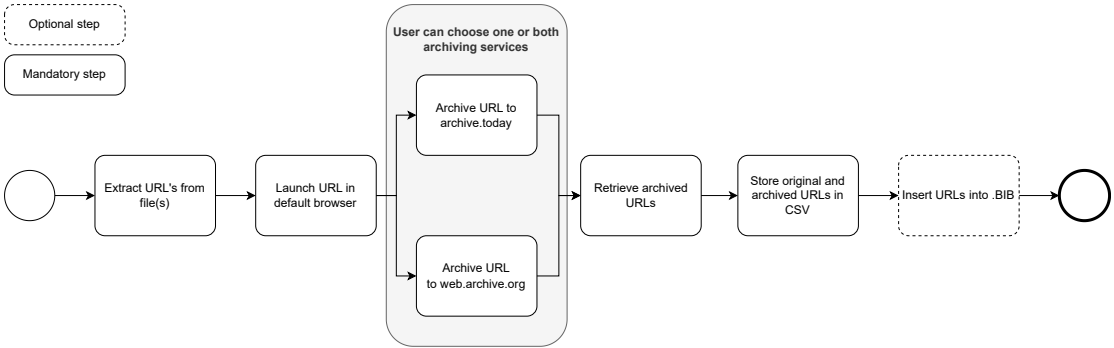
- **FolderModel:**
  - Represents a structured collection of files.
  - Can throw a ‘FolderModelException‘ during invalid operations.
- **FileModel:**
  - Represents individual file data.
  - Associated with ‘URLPair‘ objects.
  - Can throw a ‘FileModelException‘ during invalid operations.
- **URLPair:** Represents a pair of a URL and its related data.
- **URLArchiverModel:**
  - Core model handling the archiving of URLs.
  - Interacts with ‘URLExtractor‘ and ‘URLArchiver‘ utilities.
- **UserChoice:** Captures user’s decisions, possibly related to file operations or URL choices.

Note: These models ensure data integrity and consistency within our application, separating data handling from other concerns.

Process model

Workflow for URL Extraction and Archiving

- User can skip URLs, launch them, access help or quit the program at any time.



Technologies

Core Technologies

- **Java**: The primary programming language for our application.
- **LaTeX**: Used for documentation and presentation.

Supporting Technologies

- **JUnit 5**: Utilized for unit testing.
- **PDFTextStripper (PDFBox library)**: Used for extracting text from PDF documents.
- **Selenium**: Web automation tool used for tasks like inputting URLs, handling captchas, and retrieving archived URLs due to the absence of an official API from archive.today.
- **Maven**: Essential for compilation, dependency management, and building the project.

etc.

Project Management with SCRUM

Scrum-Rollen

Sprintziele

Anforderungen

Scrum Adaptionen

etc.

Blocks

Block with a title

Content.

Without title

Block types

Exampleblock

Content.

Alertblock

Content.

Example (Example environment)

Content.

section pages