Replace QPDF by PDFio as PDF manipulation library in libcupsfilters (cfFilterPDFToPDF() filter function and others)

Organization - Linux OpenPrinting
Applicant - Uddhav Phatak

Project Size - Full Size(350 Hours)

Table of Contents

Basic Information

Name: Uddhav Phatak
Major: Computer Science and Engineering
Degree: Bachelor of Technology
Year: Sophomore
Institute: Vellore Institute of Technology(VIT), Bhopal
Timezone: Indian Standard Time (UTC+05:30)

Abstract

This project focuses on standardization of the libcupsfilters code into C. As the libcupsfilters library is principally written in regular C and not C++ due to the issues which would occur due to use of C++.
But libcupsfilters still depend on one library which is written in C++, QPDF, a library for manipulation of PDF files.

We need to replace this dependency on C++, with another library PDFio. This is a PDF handling and manipulation library, as QPDF is, but it is fully written in standard, regular C, not in C++.

Why OpenPrinting for Linux?

I am only applying to Openprinting for Linux for the GSoC 2024. Being a regular user of Linux, I have always wanted to contribute to some Linux projects. Using printing softwares to print documents has fascinated me since the day I started coding. The replacement of QPDF library will make me think out of the box for something we usually don't see in my regular studies. I would love to contribute to something that makes me think unconventionally.

Deliverables

---

1. **Replacing the use of QPDF by PDFio (Required)**
   Currently, libcupsfilters uses the QPDF library, written in C++. We need to replace the QPDF library with PDFio library, which is written in C.

2. **Investigation of Suitability, and Addition of new features into PDFio Library (Required)**
   QPDF is a complex and sophisticated library with a lot of features while PDFio is a relatively new library. And we must be very careful to see whether it does not miss any important feature. Especially looking after the correct printing of filled-in PDF forms and PDF annotations.
   So part of the project will be investigation of suitability and perhaps also work to get needed features added.

Why Avoiding C++?

---

1. **Issues with Binary Compatibility in C++**
   C++ lacks a standardized Application Binary Interface (ABI), leading to compatibility issues when components are compiled with different compilers or versions. Updates or recompilations with different compilers may introduce incompatible changes, causing crashes or incorrect behavior.

2. **Difficulty in Dependency Management on Debian/Ubuntu**
   Debian and Ubuntu rely on accurate dependency information for package management, but C++'s complex build process and lack of standardized dependency management pose challenges.

   Also, automated dependency detection in these systems struggles with C++ projects' intricate dependency structures, leading to situations where certain dependencies are not properly included in the package metadata, resulting in missing dependencies or unnecessary inclusions when the package is installed on a user's system. Alternatively, it could result in unnecessary dependencies being included, bloating the package size and potentially causing conflicts or compatibility issues.

Implementation Plan/Project Timeline

Community Bonding Period May 1, 2024 - May 26, 2024

- Dive into the project's codebase, exploring its structure, architecture, and key components to gain a comprehensive understanding of how the software functions.
- Review relevant documentation, README files, and code comments to grasp the rationale behind design decisions, coding conventions, and usage patterns within the project.
- Identify critical areas of the codebase related to the project objectives, such as modules using QPDF dependencies, and analyze their implementation to anticipate potential challenges and inform your approach to the upcoming tasks.

Weeks 1-3 May 27, 2024 - June 16, 2024

- Review current usage of QPDF within libcupsfilters and identify dependencies.
- Map dependencies of QPDF with PDFio and begin function mapping from C++ to C.
- Prioritize critical areas for the transition, focusing on modules with significant reliance on QPDF.

Weeks 4-6 June 17, 2024 - July 7, 2024

- Start replacing C++ dependencies with C counterparts, ensuring core functionalities are migrated successfully.
- Conduct initial testing to validate changes and ensure the codebase remains functional after migration.

Mid-Term Evaluations July 8, 2024 - July 12, 2024

- An overview of the transition progress, including the identification of challenges and initial steps taken to address them.
- If possible, provide a demonstration of the migrated functionalities or any tangible progress achieved during the first half of the project timeline.

Weeks 7-9 July 13, 2024 - August 4, 2024

- Continue replacing C++ dependencies with C equivalents, focusing on bug fixing and addressing issues identified during testing.
- Finalize the transition and review changes made during the process to ensure completeness and correctness.

Weeks 10-11 August 5, 2024 - August 18, 2024

- Conduct comprehensive code reviews to verify successful replacement of C++ dependencies.
- Address any remaining issues or discrepancies and conduct regression testing to identify and resolve regressions.
- Prepare documentation outlining modifications and updated dependencies for future reference.

Final Evaluations August 19, 2024 - August 25, 2024

- Wrap up any remaining tasks and ensure all functionalities work as expected with PDFio.
- Prepare documentation and reports summarizing the transition process, outcomes, and recommendations for future development.

About Me

Personal Details

My name is Uddhav Phatak. I am a sophomore in the Department of Computer Science and Engineering at the Vellore Institute of Technology (VIT), Bhopal. Since my first year of college, I have been into coding and have grown my interest in exploring different fields of computer science.

I am new as a contributor to open-source.I am currently understanding the open-source community. I have been trying out various technologies ranging from web development, Computer-Networks and Machine Learning. Over the last few months, my interest has shifted towards Computer-Networks mainly the 802.11 WLAN Protocol.

In terms of my technical skills, I'm proficient in C, Python, Java, and have basic knowledge in C++, but I haven't applied it in any large-scale project yet. I am a keen learner and ready to learn anytime. That's why I feel I can compensate for this. I am also proficient with Git. I have been practicing C/C++ for a few months now in my Data Structures and Algorithms course labs.

Communication

Over the summers, I will be available to work usually around 10 am to 11 pm IST on weekdays though I can be flexible with my schedule as per the mentors' availability. I may not be available for a few days due to some family trip, but I'll inform the mentors about that beforehand. I would love to spend time with the mentors and the team learning from them over the summers.

I am comfortable with any means of communication - be it email, voice call, or video meetings, as per the mentors' convenience. For the language part, I speak English and Hindi, and Marathi.

Post-GSoC

Since this project is vast in terms of scope, some of the points mentioned may be left unimplemented. If such a case happens, I will try to complete them after GSoC. After the GSoC, I would like to keep contributing to the OpenPrinting module and explore other Linux Open Source projects.

Development Environment

I use Ubuntu 22.04 LTS as my primary operating system. I have successfully installed libcupsfilters, pdfio, CUPS and QPDF from the source code on my laptop.