Google Summer of Code 2016

Project Proposal

Zoner Draw Import Filter

Hunter Lee
hl130@duke.edu

Introduction

Problem statement

Zoner Draw 5 is a vector graphics editor with a user interface. Its main features include: tools for creating curves, rectangles, ellipses, polygons, stars, tables, barcodes, and fully editable object styles: editable pen, fill, shadow, and object operations: combinations, weld, and trim. Zoner Draw 5' s development has stopped, and now is called Zoner Photo Studio 18, but the last version (Zoner Draw 5) has been released free of charge. ZMF extension is the Zoner Callisto metafile, and the most important details of the format has already been reverse-engineered. I plan on solving this problem by creating an import for ZMF format using librevenge.

Project Facts

- Website: https://wiki.documentfoundation.org/Development/GSoC/Ideas#Implement_Zoner_Draw_import_filter
- Mentor: *David Tardon*, IRC: dtardon, mail: dtardon@redhat.com

Project Goals

List of deliverables

- Zoner Draw import filter
- Contributing to re-lab project' s OLEToy: zmf.py (for reverse engineering)
- Test results using zmf2raw.cpp

Zoner Draw 5 Format Features for implementation

- Curves
  - Lines, curves, closing curves
- Shapes (Object)
  - Rectangles, Ellipses, Triangles, Stars, Polygons
- Text
- Tables

Implementation

Basic details

My project idea stemmed from my conversations with David Tardon, a potential mentor at LibreOffice, who agreed to mentor me for this project. I have also reached out to a two-time GSoC-er for LibreOffice previously and received useful suggestions and advice. I have built LibreOffice from its Git repository and have been working on an easy hack (https://bugs.documentfoundation.org/show_bug.cgi?id=32531).

What I would like to achieve over the duration of Google Summer of Code is as follows. My project will focus on creating a Zoner Draw import for LibreOffice, creating a separate library using librevenge (document liberation importer). librevenge is a base library for writing document import filters. It has interfaces for text documents, presentations, vector drawings and spreadsheets. To give an example, libpagemaker and libmspub were two recent libraries started as GSoC projects based on librevenge.

Based on my discussions with my potential mentor, I plan on providing a sample to familiarize myself with librevenge by starting a skeleton of a new library created by David. I will start by working with rudimentary functionalities such as format detection. This would replace my easy hack on LibreOffice codebase, but I plan on continuing fiddling around with the easy hack I have started on the side to understand what LibreOffice codebase looks like.

Overview

Project outline

- librevenge: a set of interfaces for typical office file formats that also conveniently supports creation of a skeleton of a new project. A skeleton code is available and will be worked on for an easy task like format detection.
    1. Zoner Draw Skeleton provided (*libzmf*)
    2. librevenge uses CppUnit (Unit testing library), so I plan on using it for unit tests
    3. C++ Boost library will possibly be used for number handling, string handling, serialization, and streams.
- Reverse engineering: OLEToy—Python scripts to parse and view contents of various binary file formats.
    1. Identifying the file format
        - I have used a hex-editor to look at ZMF files. I plan on continuing this paying special attention to fillers and zeros
        - Determine endianness (big) and identifying files using the method above
    2. Major Zoner Draw 5 objects have already been reverse-engineered. The details can be found here (https://github.com/renyxa/re-lab/blob/master/oletoy/zmf.py)
    3. Study _parse_contents and _parse_objects in zmf.py
    4. OLEToy can help in rev-engineering, reducing the need to use serialization and alignment in C++
- Testing
    1. Test using custom perl scripts using the foo2raw tool
        - *zmf2raw.cpp* is included in *libzmf* skeleton
        - Runs librevenge calls and the printout will be compared using diff

Timeline

Key Dates

Timeline	Tasks
Community Bonding/Prep/Easy Hack	Work on the skeleton library, starting with format detection
Week of 23MAY2016	GSoC starts
Week of 20JUN2016	Midterm evaluation preparation
Week of 27JUN2016	Midterm evaluation submission
Week of 15AUG2016	Final Wrap-up

- I do not have any vacation plans or prolonged breaks planned; however, I might be travelling to my parents' over the summer, but that will be communicated as it arises (at most 24-48 hours of flight time).

From my discussion with my mentor, because this project involves reverse engineering, setting exact dates are rather unrealistic and difficult. I am presenting a rough stage-based development plan as follows:

1. **Planning** (by end of June): This will involve getting deeper into reverse engineering and reading other librevenge-based libraries (libwpg, libmspub, etc) to get a better idea of how I should design and write libzmf.
2. **Development** (by end of July/early August): This will involve implementing a small piece of functionality (a feature or a subfeature) and soliciting feedback from my mentor (David) to ensure that it meets expectations.
3. **Wrap-up** (by mid—August): Completion of features that have been worked on. Assess the difficulty of reverse engineering for other features. Project roadmap.

About me

Why LibreOffice should choose me (resume attached)

- NAME: Hunter Lee (Hee Boung Lee)
- E-MAIL: narendly@gmail.com (personal), hl130@duke.edu (school)
- IRC Nickname (freenode #libreoffice-dev): hlee
- GitHub: @narendly
- TEL: 919-317-2113
- Timezone: EST (UTC -5), UTC+09:00

This is my first time participating in Google Summer of Code and open-source development. I spoke with a previous 2-time GSoCer for LibreOffice and became interested in the opportunity. I am currently majoring in Electrical and Computer Engineering and Computer Science at Duke University, and my academic coursework gave me a good exposure to software development. I have started a few years ago playing around with algorithmic problems in C/C++, just a few of which I keep on my GitHub page.

Zoner Draw 5 is a nice program that lets users create documents containing vector-based diagrams and is useful for creating unique and distinct images with effects. Understanding of reverse-engineering serves as a basis for this project, and from my experience with working with operating system concepts, stack-smashing, and MIPS Assembly, I believe I will be a qualified contributor in a meaningful way.

Documentation Liberation Project

The DLP (Document *Liberation* Project) had a personal appeal because I have been exposed to situations where I had difficulty reading and writing documents due to format issues (for example, HWP format in Linux-based OSes). As part of the DLP, this project will benefit LibreOffice a great deal as it will draw more users interested in images using vector graphics and add further meaning to the project itself.