The Journal of Open Source Software

General binary file parser.

Jeroen F.J. Laros[1]

1 Leiden University Medical Center,

DOI: 10.21105/joss.00766

Software

- Review
- Repository
- Archive

Submitted: 03 June 2018
Published: 06 June 2018

Summary

To enable interoperability for (proprietary) binary file formats while maintaining compatibility with the original software, a lot of effort needs to be put in reverse engineering of the file format and the subsequent programming of a dedicated parser and writer. Here we introduce a library that aims to keep this effort to a minimum by providing a framework that enables a programmer to record the knowledge gained in the reverse engineering process in a structured way for it to be used directly in a parser, a writer and as documentation.

General binary file parsing and writing is implemented by interpretation of human readable documentation of the file structure and data types. Basic types like variable length strings, maps and bit fields (flags), as well as elementary data types provided by the struct library (Python Software Foundation 2001–2018, Gindi (2013)) are supported and other data types are easily added. Apart from basic types, nested structures and various kinds of iterators are supported to accommodate for complicated file formats. Numerous character encodings are supported via the iconv library (Shtuchkin 2011).

Since all operations needed for parsing a binary file can be reversed, fully functional binary editing is possible using this library. A binary file can be converted to a serialised dictionary representation, edited and be converted back to its binary form.

We have made two implementations of this library: one in Python and one in JavaScript. We chose YAML (Simonov 2011–2015, Puzrin (2011–2015)) as our preferred serialised dictionary format, but other serialisation formats (JSON for example) can be used too.

References

Gindi, Daniel Cohen. 2013. “Python-Struct - Packs/Unpacks/Measures Structs According to Python’s Struct Format.” https://github.com/danielgindi/node-python-struct.

Puzrin, Vitaly. 2011–2015. “JS-YAML - YAML 1.2 Parser / Writer for JavaScript.” https://github.com/nodeca/js-yaml.

Python Software Foundation. 2001–2018. “Struct — Interpret Strings as Packed Binary Data.” https://docs.python.org/2/library/struct.html.

Shtuchkin, Alexander. 2011. “Iconv-Lite - Pure JS Character Encoding Conversion.” https://github.com/ashtuchkin/iconv-lite.

Simonov, Kirill. 2011–2015. “PyYAML - the Next Generation YAML Parser and Emitter for Python.” https://github.com/yaml/pyyaml.