

# JOSS\_bibliographies

July 10, 2019

## 1 Background

Publications of the Journal of Open Source Software (JOSS<sup>1</sup>) are equally available as .html and .pdf, altogether with their corresponding .xml files, on a separate public repository<sup>2</sup>. This not only offers an alternative access to the publications *at once*, equally, it is possible to retrieve their bibliographic data rapidly for further processing in literature reference programs like zotero<sup>3</sup>

## 2 Technique deployed

The uncompressed .zip archive is cloned onto a machine running Python3. The branch was chosen both because of the anticipated retirement of Python2, as well for the better support of Unicode characters. Both scripts run from the CLI, and expect the .zip archive in the same folder, e.g.

```
python bibtool_Joss_pocket.py joss-papers-master.zip
```

to retrieve the necessary raw data, which is either the .html or .crossfile.xml – but not the substantial larger .pdf.

The smaller script, joss-papers-master.zip, retrieves the small «cite as» block provided on each .html of the corresponding publication, in either .txt or .csv format. The larger bibtool\_Joss\_bibtex.py attempts to retrieve *basic* bibliographic information and creation of a bibtex file (.bib) understood by many literature reference programs. In addition, a less verbose .csv (separator: semicolon) is created, offering further processing with any editor, or spreadsheet.

After running these scripts, the scripts leave you with the original .zip archive (still containing the .html and .pdf about the publications) and the extracted literature data. Intermediate files are deleted automatically.

## 3 Copyright

(c) 2019 Norwid Behrnd, GPLv3.

---

<sup>1</sup><https://joss.theoj.org/>

<sup>2</sup><https://github.com/openjournals/joss-papers>

<sup>3</sup><http://www.zotero.org>