PubMed's Scanning:
Raw data from *
Number of papers (n = 42,397)

Extract from abstract

Link found in Abstract
(n = 7,287)

Auto-check:
Using script from **

Active link (<400):
(n = 5,267)

Non-active link

Broken status (≥400):
(n = 929)

Time-out (status = -1)
(n = 1,091)

Manual Check:
0% false positive

No link in abstract

Extract link from Body of Paper
(n = 30,203)
Auto-check: Using script from **

Active link (<400)
(n = 21,562)

Non-active link

Broken status (≥400)
(n = 5,137)

Time-out (status = -1)
(n = 3,504)

- .tar.gz files that begin with "article" on ftp://ftp.ncbi.nlm.nih.gov/pub/pmc/
- ** script found on https://github.com/smangul1/good.software