|  |
| --- |
| Table 1 Pseudocode to download and pre-process the Portable Document Format (PDF) of each article collected for this study. Readers are referred to the cited GitHub repository for full implementation using Python. |
| Algorithm 1: Pseudocode to add each article PDF to a manipulable dataset |
|  |
| input: *X is the journal archive website* (e.g., https://www.ncbi.nlm.nih.gov/pmc/journals/309/) |
|  |
| output: *M is the data frame with the features of the journal article, the year, binned years, cleaned data* |
| top:  Initialize U 🡨 list of the unique urls for a given journal  Initialize Y 🡨 Unique folder for article    Visit the archive website and retrieve all unique journal issue link  For each unique jounal issue link |
| loop:  for ui ∈ U do  use Selenium and Beautiful Soup via bot to visit the unique url  download the article  initialize G 🡨 list of pages in article  for gn ∈ G do  initialize D 🡨 txt file  resized first page 🡨 page without border of article  all text 🡨 all pages converted to text in the article  append all data extracted as text to the text file.  add text document to Y  for yi ∈ Y do  add file name to M  add text to data frame  return M |
|  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Year** | **Journal** | **Word Count** | **Cleaned Data** | **Decade** |
| 1996 | Journal of the Experimental Analysis of Behavior | 39116 | journal identify naming basic unit verbal behavior describe | 1990-1999 |
| 1958 | Journal of the Experimental Analysis of Behavior | 36476 | review critically analyzes experimental data relevant | 1950-1959 |
| 1968 | Journal of the Experimental Analysis of Behavior | 32592 | journal schedule reinforcement maintained pigeon key | 1960-1969 |
| 1996 | Journal of the Experimental Analysis of Behavior | 31170 | journal introduction next paper comprise set gin theoretical | 1990-1999 |
| 1981 | Journal of the Experimental Analysis of Behavior | 28415 | journal year paper selective review method problem finding | 1980-1989 |
| 1999 | Journal of the Experimental Analysis of Behavior | 28341 | journal propose fundamental unit behavior concurrent | 1990-1999 |
| 1983 | Journal of the Experimental Analysis of Behavior | 28112 | journal research multiple schedule interaction reviewed | 1980-1989 |
| 1962 | Journal of the Experimental Analysis of Behavior | 27850 | although theorist may found frequent controversy experiment | 1960-1969 |
| 1999 | Journal of the Experimental Analysis of Behavior | 27090 | book review published year ago drew attention similarity fun | 1990-1999 |
| 2001 | The Behavior Analyst | 24397 | spring intrinsic motivation major concern psychology | 2000-2009 |
| 2020 | The Behavior Analyst | 23815 | perspective http doi orgs check update da online pairing | 2020-2021 |
| 2018 | Journal of Applied Behavior Analysis | 23772 | journal present paper provides integrative review research | 2010-2019 |
| 2009 | Journal of the Experimental Analysis of Behavior | 23594 | journal last decade researcher proposed large number | 2000-2009 |
| 1978 | Journal of the Experimental Analysis of Behavior | 23055 | journal model conditional discrimination learning pigeon | 1970-1979 |
| 2005 | Journal of the Experimental Analysis of Behavior | 22995 | journal equilibrium phenomenon matching behavior traditional | 2000-2009 |
|  |  |  |  |  |

**Table 2.** Example rows of the analytic dataframe created by downloading all .pdfs, converting each article to text, pre-processing and cleaning the text, and appending each article as a single row to the aggregate dataframe.