

# HTRC Derived Datasets

## Introduction

HTRC releases research datasets to facilitate text analysis using the HathiTrust Digital Library. While copyright-protected texts are not available for download from HathiTrust, fruitful research can still be performed on the basis of non-consumptive analysis of features extracted from full text. These features include volume-level metadata, page-level metadata, part-of-speech-tagged tokens, and token counts. Additionally, HTRC has partnered with advanced researchers to release a derived dataset, *Word Frequencies in English-Language Literature, 1700-1922*.

## Getting Started

[Downloading Extracted Features](#)

[Use Cases and Examples](#)

[Extracted Features in the Wild](#)

## Extracted Features Dataset

[Documentation](#)

[Get the data](#)

## Word Frequencies in English-Language Literature, 1700-1922

[Documentation](#)

[Get the data](#)

## Extracted Features Dataset [v.0.2]

**NOTE:** this dataset has been superseded by Extracted Features Dataset [v.1.0], above.

[Documentation](#)

[Get the data](#)