University of Kentucky

UKnowledge

2-14-2024

What's Gobbling this Hard Drive? Deduplication, Advanced File Search, and File Characterization Beyond Your Wildest Dreams with TreeSize For Windows

Andrew McDonnell
University of Kentucky, mcdonnell@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/ri_day_2024

Part of the Archival Science Commons

Right click to open a feedback form in a new tab to let us know how this document benefits you.
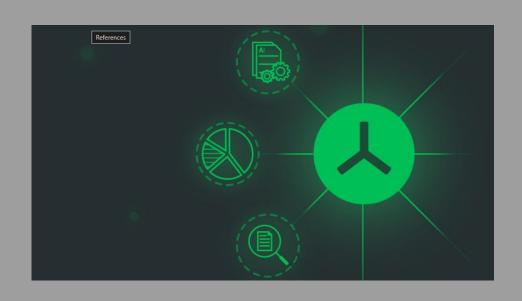
TreeSize Demo:
What's Gobbling this Hard Drive?

Deduplication, advanced file search, and file characterization for archival processing and better living!

What is TreeSize?

Primary Sales Pitch:
Analyze, manage,
and tidy storage systems

- Windows OS only
- Can examine drives, folders, SharePoint, Outlook, S3, + more
- Product of Jam Software out of Trier, Germany.

Value to Born-Digital Processing: Immense

Deduplication

Value to Archival Processing: Immense
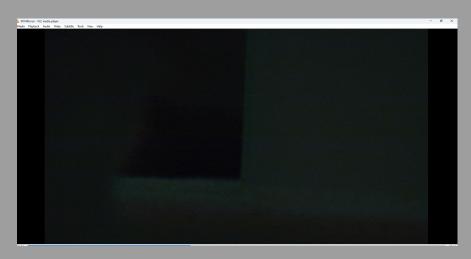
Deduplication

Example:
Kentucky Quilt Project
Pre-processing:
146,665 files

TreeSize analysis:
~60,000 duplicate files
(some recurred 10x
across multiple
directories)

Value to Archival Processing: Immense

Large File Characterization

360 MB: 1:11 silent clip from a dark room

Example:
Kentucky Quilt Project

Final Cut scratch disk:
2.89 TB, 8,500 files

Value to Archival Processing: Immense

- PII identification
- Bulk renaming
- Difficult format ID
- File/folder search
- Illegal characters
- Batch exports
- File manifests

Example: Prichard Committee for Academic Excellence

Obsolete + Unsustainable formats
10,363 files across 1,316 folders

A Demonstration

Podge, I wish to sample these wares...

Free 30-day Trial: www.jam-software.com/treesize

- License is perpetual
  (rather than ongoing subscription)
- Libraries currently hold 8 licenses. Potential for more if need is evident.

Questions?