



Metadata-Focused File Management Systems



Jackson Tenor
UMM CSci Senior Seminar
Spring 2020

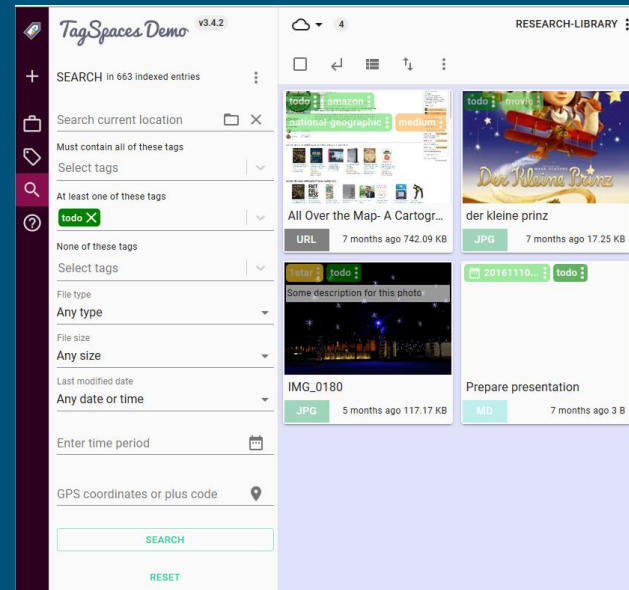
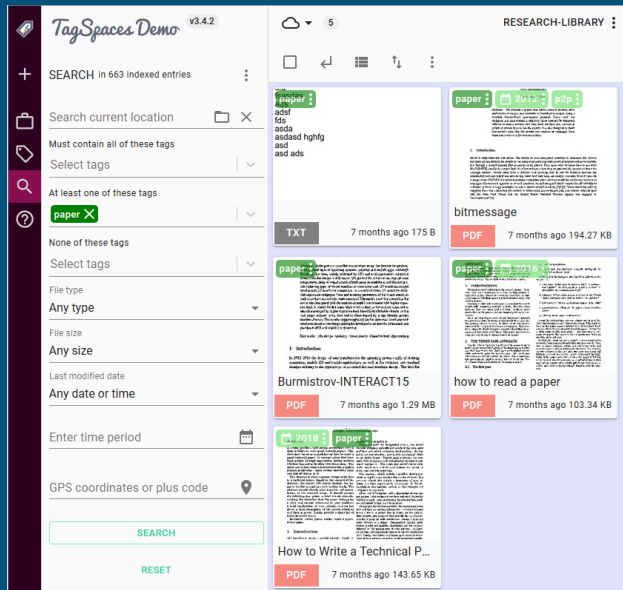


Overview - Problem

- The size of file collections has increased rapidly
- The hardware we store data on has similarly changed
- Existing file systems have not adapted to these changes
- This has resulted in large, obtuse, difficult to use file systems

Overview - Solution

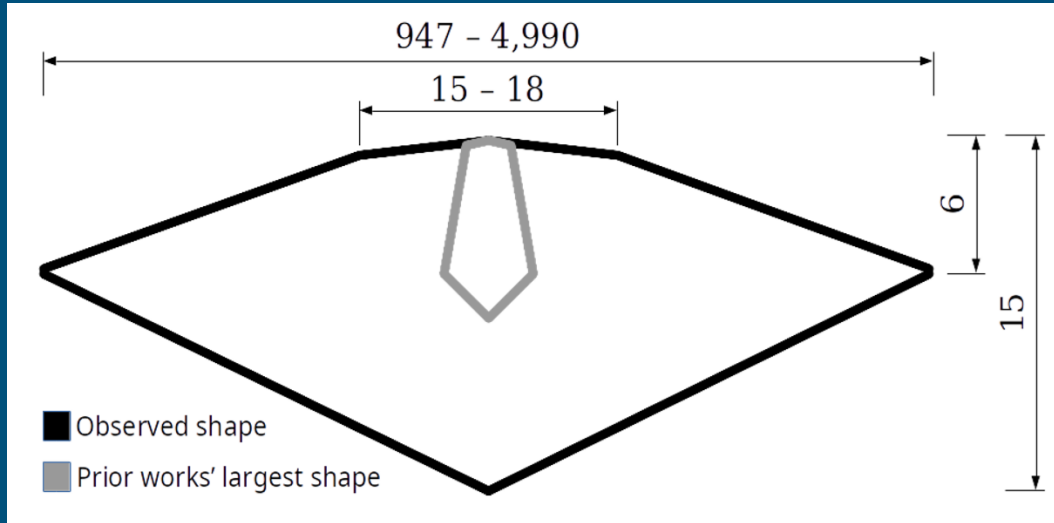
- Metadata-focused file management helps manage modern file collections



1. Background
2. Exploring Metadata-Focused
File Management
3. Tree Tags
4. Conclusions

Background

Changes in File Collection Size



A comparison of modern file systems to prior works (2005-2010)

File System Standards

- POSIX
- Nearly universal
- Dictate how operating systems should work
- Any new solutions must consider these

File System Standards

- Commands

- opendir
- open

- Paths

- Directory: /Images/2019
- File: /Images/2018/IMG01.png

- Images

- 2018
 - IMG01.png
 - IMG02.png
 - IMG03.png
- 2019
 - IMG04.png
 - IMG05.png
 - IMG06.png

Existing Metadata - File System

- Examples

- Creation Date
- Owner

```
total 164
-rw-rw-r--. 1 jacksontenor jacksontenor 39097 Apr 10 08:37 code.png
-rw-rw-r--. 1 jacksontenor jacksontenor 43901 Apr 10 08:41 lamp.png
-rw-rw-r--. 1 jacksontenor jacksontenor 80671 Apr 15 16:53 parachute.jpg
```

- Issues

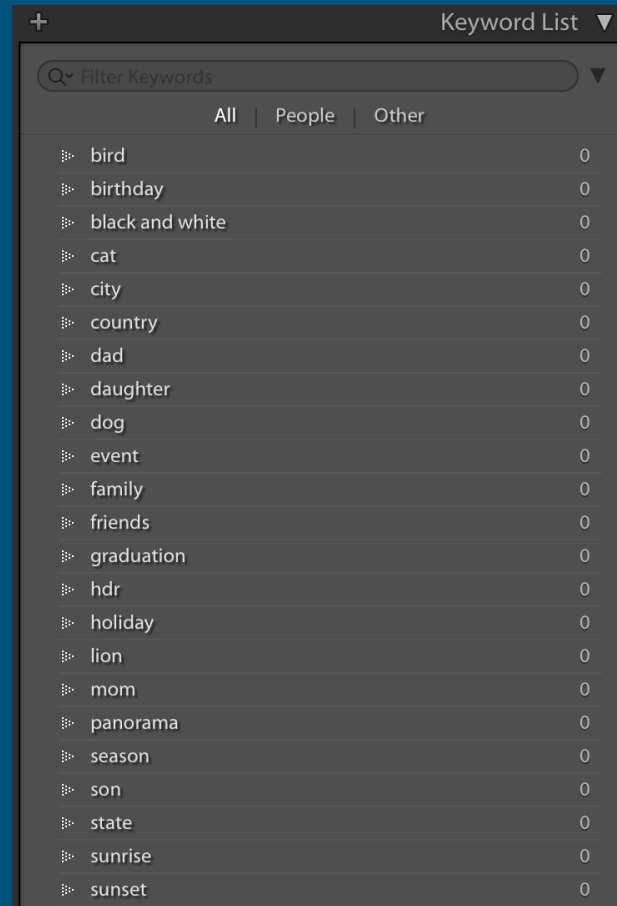
- Not user defined or oriented
- Restricted to specific data

Existing Metadata - EXIF

- Examples
 - Camera Equipment
 - Photographer
 - Copyright
- Issues
 - Not universal to all file formats
 - Requires read of file
 - Not user defined

User Defined Metadata

- Ensures phrasing familiar to users
- Information users care about
- Most common types
 - Tagging
 - Key-Value pairs
 - Hybrid



The screenshot shows a 'Keyword List' window with a search bar and three tabs: 'All', 'People', and 'Other'. The 'All' tab is selected, displaying a list of 20 keywords with their respective counts. Each keyword is preceded by a small icon of three dots.

Keyword List ▼	
Q Filter Keywords ▼	
All	People Other
bird	0
birthday	0
black and white	0
cat	0
city	0
country	0
dad	0
daughter	0
dog	0
event	0
family	0
friends	0
graduation	0
hdr	0
holiday	0
lion	0
mom	0
panorama	0
season	0
son	0
state	0
sunrise	0
sunset	0

Types of File Management Systems

- Application Based (Adobe Lightroom)
 - Limited by platform
- Cloud Based (Google Photos)
 - Limited by platform
 - Limited in ownership
- Desktop Search
 - No understanding of file content
 - No metadata added
- Native File System
 - Biggest change

Exploring Metadata- Focused File Management

Relationship with Hierarchy

- Pros
 - Familiar
 - Backwards compatible
 - Can utilize existing tools
- Cons
 - May restrict design space
 - May cause metadata to be ignored

File Level Metadata

- Tags
 - Key value pairs
-
- Free form strings vs defined possible values
 - Single value vs multi value
 - How it is applied to collections, if hierarchy is used
 - How are queries on the metadata structured

Updating Metadata

- Existing queries can read, write, create, or delete files
- New queries must be available to touch the metadata rather than the content of the file
- May be applied to one file or many

Adding Collections

- How can users store views of data
- Stored or dynamically generated
- Updated automatically or manually

Tree Tags

Tree Tags

- Adds tags to hierarchical folders and files
- Add support for metadata additions and querying
- Designed to bridge the gap between traditional and new file systems

Tree Tags - Querying the File System

- Images
 - 2018
 - IMG01: Ryan
 - IMG02: Kate
 - IMG03: Ryan, Kate
 - IMG04: Ryan, Kate
 - IMG05: Jon
 - 2019
 - IMG06: Kate
 - IMG07: Kate, Jon
 - IMG08: Ryan, Kate, Jon

FileName: Tags

Traditional commands with metadata in the path

- /Images/2018/Ryan
 - Returns images 1, 3, 4
- /Images/2018/Ryan [OR] Kate
 - Returns images 1-4
- /Images/2018 [OR] 2019/Ryan [AND] Kate
 - Returns images 3, 4, 8

Tree Tags - Additional Operations

- Metadata-focused commands to add and remove tags
- Metadata operations work on both collections and files
- Collection tags propagate downwards to their children

Tree Tags - Remaining Hurdles

- Implementation
 - Design of data structures, algorithms
 - Usable prototype
- Benchmark comparison

Conclusion

- Computer hardware has changed
 - The scope of file collections has increased
-
- Metadata-focused file management solves many existing issues
 - Implementation decisions remain

Acknowledgements

Thanks to Nic McPhee and KK Lamberty for their advising and guidance.

Thanks to Sydney Richards for her feedback.

Questions

References

Exif. Wikipedia.

Adobe. Adobe photoshop lightroom classic user guide.

N. Albadri, R. Watson, and S. Dekeyser. Treetags:Bringing tags to the hierarchical file system. In Proceedings of the Australasian Computer Science Week Multiconference, ACSW '16, New York, NY, USA, 2016. Association for Computing Machinery.

J. D. Dinneen, C.-A. Julien, and I. Frissen. The scale and structure of personal file collections. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, New York, NY, USA, 2019. Association for Computing Machinery.

A. W. Leung, M. Shao, T. Bisson, S. Pasupathy, and E. L. Miller. High-performance metadata indexing and search in petascale data storage systems. Journal of Physics: Conference Series, 125:012069, jul 2008.

S. Ma and S. Wiedenbeck. File management with hierarchical folders and tags. In CHI '09 Extended Abstracts on Human Factors in Computing Systems, CHI EA '09, page 3745–3750, New York, NY, USA, 2009. Association for Computing Machinery.

Y. Son, H. Han, and H. Y. Yeom. Optimizing file systems for fast storage devices. In Proceedings of the 8th ACM International Systems and Storage Conference, SYSTOR '15, New York, NY, USA, 2015. Association for Computing Machinery.

L. Todd. Posix file system basics, 2017.

R. Watson, S. Dekeyser, and N. Albadri. Exploring the design space of metadata-focused file management systems. In Proceedings of the Australasian Computer Science Week Multiconference, ACSW '17, New York, NY, USA, 2017. Association for Computing Machinery.