

Extracting Data: Text from Photos

Emily Una Weirich

Center for Creative Photography, University of Arizona

Artifact type: Workshop

Developed for: Undergraduate or graduate courses, or professional development workshops

Audience: Photographers, humanities students, LIS professionals

Time required: 1 90-minute workshop session

Method and tool: User interface and Command Prompt; metadata extraction and ExifTool

Description:

This workshop would be useful for those who want to know what textual information already exists about a collection of photos by using ExifTool (<https://www.sno.phy.queensu.ca/~phil/exiftool/>). While this exercise would provide at least some metadata for any digital photographs, it would be a richer workshop when working with photos that have both automatically and human generated metadata. For example, I have done this exercise with a collection of photographs by a photojournalist who created captions, headlines, and keywords for their photographs. I would recommend providing participants with a folder of pre-selected images that you know in advance would work for the workshop.

The workshop is generally for those who have already been introduced to the command line and are ready to practice using it. In some of these cases, it might be more beneficial to have students focus their metadata evaluation on specific fields rather than allowing them to choose which ones they find interesting. The workshop could be scaffolded into larger course arcs in order to gain skills related to digital humanities tools used to create digital exhibitions, mapping projects, or conduct text analysis; to teach skills related to processing digital archives or digital preservation; and to generally help students in developing critical thinking and information analysis skills necessary for digital information literacy.

As written, the purpose of the workshop is to give students a simple command line exercise and introduce them to embedded photographic metadata. At the conclusion of the workshop, participants should be more comfortable using the command line for simple everyday tasks such as navigation and creating folders, they should have a basic understanding of how to use ExifTool, and they will end up with a data set about a collection of photographs that could be used in future assignments or workshops.

Supporting materials:

Workshop plan

Extracting Data: Text from Photos

Your Name: Emily Una Weirich

Your institution: Center for Creative Photography, University of Arizona

In this workshop, you will have practice your command line skills while learning to use ExifTool to extract textual metadata from digital photographs. Note that this workshop is suitable for those using Windows and Mac operating systems; commands are written side-by-side when necessary.

Setup - PC

Install ExifTool and setup the folders and files you will use in this workshop.

1. Download the latest version of ExifTool:
<https://www.sno.phy.queensu.ca/~phil/exiftool/index.html>
 - a. Download the "Windows Executable" file and save the file to your "Downloads" folder
 - b. Follow the download instructions:
<https://www.sno.phy.queensu.ca/~phil/exiftool/install.html>
2. Open Command Prompt
3. Navigate to the folder with the files you just extracted and confirm that the .exe file is there:
 - a. `$ cd [downloads]`
 - b. `$ dir`
 - c. `$ cd exiftool-[version]`
 - d. `$ dir`
4. Rename the file and check your work:
 - a. `$ ren exiftool(-k).exe exiftool.exe`
 - b. `$ dir`
5. Copy that file to the Windows folder on your computer

Setup - Mac

Install ExifTool and setup the folders and files you will use in this workshop.

1. Download the latest version of ExifTool:
<https://www.sno.phy.queensu.ca/~phil/exiftool/index.html>
2. Follow the installation instructions:
<https://www.sno.phy.queensu.ca/~phil/exiftool/install.html#OSX>

Start it up!

1. Head to **PC** Command Prompt or **Mac** Terminal
 - a. `$ exiftool`
 - b. Wait for it...
 - c. Eventually, the ExifTool manual will load. ExifTool is now running! Scroll through the ExifTool manual by pressing the space bar.
2. Navigate to your desktop and make a folder to contain the files for this workshop
 - a. **PC** `$ cd .. / Mac $ cd -`
 - b. `$ cd Desktop`
 - c. `$ mkdir PhotoData`
3. Copy the "photos" folder containing a selection of photos to the PhotoData folder

Try it out

You will get to try using ExifTool to extract metadata from a single file and refine the data that you are extracting.

1. Head back to **PC** Command Prompt or **Mac** Terminal

2. Navigate to the photos folder
 - a. `$ cd photos`
3. Take a look at the photo filenames in that folder and take note of the file name of a file that you want to look at first
 - a. **PC** `$ dir` / **Mac** `$ ls`
4. Extract some metadata!
 - a. `$ exiftool fileName.jpg`
 - Note: When typing in file names, be sure to duplicate capitalization and file extensions.
 - b. Give it a few seconds to process your request
 - c. Take a few minutes to look at the data. And consider the following in a small group:
 - What looks interesting?
 - Who might find the different types of data listed useful?
 - Which fields do you think were created automatically in the camera or by a computer program, and which do you think were entered later by a person?
 - What do you think the data could tell you?
 - Are there any things that you have questions about or don't seem right?
 - Which fields would you be interested in extracting from across all of the images in this set? Please choose at least three.
5. Extract the metadata for the fields you are most interested in by entering a string of commands for ExifTool
 - a. `$ exiftool -fieldOne -fieldTwo -fieldThree filename.jpg`
 - b. Review what ExifTool has produced for you and update your command string as needed

Scale it up

Let's run ExifTool on a whole folder of images and create a spreadsheet with the extracted metadata.

1. Navigate back to the main project folder
 - a. **PC** `$ cd ..` / **Mac** `$ cd -`
 - Note: From here we can point ExifTool to an entire folder instead of just one file within that folder
2. Create a CSV file with metadata extracted from the three fields you identified earlier for all of the photos in the "photos" folder
 - a. `$ exiftool -csv -fieldOne -fieldTwo -fieldThree photos>photoMetadata1.csv`
3. Open this new file
 - a. **PC** `$ photoMetadata1.csv` / **Mac** `$ open photoMetadata1.csv`

Consider some possibilities

Now that you have your data set, take a few minutes to consider the following with your group:

- If you were to build a digital humanities project using this data set and the associated images, what would you want to do?
- Who would your audience be and why would they be interested in spending time with your project?
- What tools could you use?
- What additional information would you need?
- Now that you've spent some time with embedded photographic metadata, which fields would you like to use in your own photographic practice?
- What problems or inconsistencies can you find in your data set?
- What potential difficulties might these problems cause for photographers? For archivists? For curators? For researchers and students?

Additional resources

- ExifTool: <https://www.sno.phy.queensu.ca/~phil/exiftool/>
- AVP, *Exiftool Tutorial Series*, by Kara Van Malssen: <https://www.weareavp.com/exiftool-tutorial-series/>

Acknowledgement

This workshop plan was created as a component of “[Textual Data and Digital Texts in the Undergraduate Classroom](#),” a 2018-2019 NEH Institute for Advanced Topics in the Digital Humanities, which was organized by Lauren Coats and Emily McGinn.