Using Google Cloud Vision and Chrome Extensions for Improved Metadata

Aaron O’Donovan
Special Collections Supervisor
Metadata is Big Data

• Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional resources [https://en.wikipedia.org/wiki/Big_data](https://en.wikipedia.org/wiki/Big_data)
Common Problems with Creating Metadata

- Incomplete Information
- Quality is poor
- Little context to connect it to something greater
Big Data is Big

• When surface metadata is scarce it makes sifting through the endless supply of data online very hard because of the 3 Vs

• Volume (880 billion images a year), Variety (Unlimited ways of uploading on many different sites and platforms), and Velocity (speed at which images are put online, even the best website crawlers have a hard time keeping up)
Refining and Enhancing Metadata

- Open source tools are available to help get a signal through all the noise
- Artificial Intelligence and Machine Learning can be used for more precise results and solutions that took hours or days of work
- What can be done?
  - Facial Recognition
  - Optical Character Recognition (OCR) and Handwriting Recognition (HWR)
  - Building Identification
  - Colorization of photos
A Case Study

• The Call and Post Collection loaned by the King Arts Complex

• African American paper from the 1960s-1990s

• Tens of thousands of photos with various levels of metadata

• Mostly dates but not much else on the earlier photos

• Microfilm not a help since it was not filmed in the early years of its existence
Tools for improving our metadata

• **Reveye** Reverse Image Search Chrome Extension
• Perform a search by image. Choose between the image search engines Google, Bing, Yandex, TinEye and Baidu
• Bing by far the best but they all have their advantages and disadvantages and use cases
Bing Reverse Image Search
Bing Reverse Image Search
CROCKETT Monte J. Crockett, 73, went to be with the Lord on June 13, 2012 at his residence. He was born in Talcott, W. Va. on July 14, 1958 to Arelious and Danese Crockett. Monte graduated from Lincoln High School in Hinton, W.Va and Highland University in New Mexico. He was a member of Mt. Olivet Baptist Church for many years. He played for the Buffalo Bills as a tight end in the early 1960's. He worked many years in the youth services field and then retired as the Deputy Director of Franklin County Juvenile and Domestic Court system. He was preceded in death by his father Arelious Crockett, sisters Carol Chatman and Sharon Archie. He is survived by his loving wife, Peggy A. Crockett; son, Jared (Shayne) Crockett; daughter, Kyra (Giovanni) Crockett-Hodge; mother, Erma D. Crockett; brothers, Robert, Darryl, Kenneth (Jennie) Crockett; sisters, Brenda Smith e’ Ineas, Jacqueline Crockett; grandchildren, Jasymn and Myles Crockett, Dallas and Chase Hodge; brother-in-law, Edward (Stephanie) Johnson; and a host of nieces, nephews, other relatives and close friends. The celebration of life service will take place on Wednesday, June 20, 2012, at 1 p.m., at Mt. Olivet Baptist Church, 428 E. Main St. Rev.
Bing Reverse Image Search
Caveats

• Search engines prefer the subject looking at the camera
• Profile shots don’t come up with great results
• Need high resolution images
• Works great for online but doesn’t have to be
• You can research the image before putting it online
• Subject or topic needs to have some internet presence
• Most likely not going to find that random person on a cabinet card, although you may get some better metadata to describe it
Tools for improving our metadata

• **Google Cloud Vision** Chrome Extension
• Using this for OCR and HWR
• API is more powerful and can perform other tasks
• [https://cloud.google.com/vision](https://cloud.google.com/vision)
Google Cloud Vision
Google Cloud Vision

• Copies to Clipboard
• Full resolution images give best results
• Available through the free trial API
• https://cloud.google.com/vision/docs/drag-and-drop?hl=bg
Google Cloud Vision
Google Cloud Vision
Caveats

- Maximum Image size is 4 MB for free API
- GIGO: Garbage In, Garbage Out
- TITO: Treasure In, Treasure Out
- Cursive: Poor/Good Print: Better Typed: Best
- Not out of date OCR software, keeps improving but it isn’t perfect by any means
- Works well for small batches, need the API for large batches, costs involved in that case
Tools for improving our metadata

• Google Reverse Image Search with added search qualifiers
• Great for information you already know but is incomplete
• A building with an address but an unknown location or street
Google Reverse Image Search

House in North-University Area

Search this record

house

2 Results found
Google Reverse Image Search

Possible related search: cottage

Cottage - Wikipedia
A cottage is, typically, a small house. It may carry the connotation of being an old or old-fashioned building. In modern usage, a cottage is usually a modest, often cosy dwelling, typically in a rural or semi-rural location. Wikipedia

Cottage houses

Visually similar images
Google Reverse Image Search

Results for **1606 columbus ohio**

www.zillow.com › 43205 › Franklin Park
**1606 E Rich St, Columbus, OH 43205 | MLS #219046068 ...**
Tax abatement in process making this home more affordable. Read more. Listing Agent: Misty Linn. Core Ohio, Inc ...

www.zillow.com › 43235 › Worthington Green
**1606 Fallhaven Dr, Columbus, OH 43235 | Zillow**
The property 1606 Fallhaven Dr, Columbus, OH 43235 is currently not for sale on Zillow. View details, sales history and Zestimate data for this property on ...

Visually similar images
Value Added Tools

• Colorization Sites or Apps
• Good for buildings, sometimes colors are pretty accurate and give clues to color of homes
• Many options, all slightly different
• Colourise.sg, Algorithmia, Adobe Scribbler, Remini, Topaz
• It’s fun and has future potential
Colorize Images
Colorize Images

Photo Colorization Before and After

Drag the purple line to reveal the before and after.
Thank you!

• Questions?
• Contact info:

aodonovan@columbuslibrary.org