Search on Commons

Structured Data Across Wikimedia (SDAW)

Product Design Strategy

Daisy Chen
June 2020

WIKIMEDIA FOUNDATION

Introduction

Problem statement & overview

Key Takeaways

Detailed Findings

- 1) Image search, generally
- 2) Current Commons Search
- 3) Prototype search
- 4) Static search mock-up
- 5) Preferred search interface

Recommendations and next steps

WIKIMEDIA FOUNDATION

Problem Statement & Overview

One of the stated goals for the **Structured Data across Wikimedia (SDAW)** team is to reimagine and improve search on Commons.

Following the creation of a prototype Commons search and a static mock of an even more detailed and modern Commons search look, this research project aims to address the usability of these possibilities both in relation to experience of the current Commons search, and in the context of a range of Commons user personas and their search behaviors.

> *When it becomes easier to search Wikimedia Commons - Wikimedia contributors can more effectively **illustrate** Wikimedia projects such as Wikipedia.*
>
> *- About Structured Data on Commons*

Key Takeaways

10 casual image search Usertesting sessions

- Google Images search provides the default image searching paradigm for all participants
- None had previously used Commons search
- From current Commons search, to the prototype search, to the static search mock-up; on the whole, on the whole participants felt the progression represented incremental improvement. Static mock-up received near universal top marks.
- The concept chips were readily apparent to this cohort, but were well-received with all users recommending inclusion
- Poor search relevancy, tangential to this project's goals, was a motif

5 Commons users moderated sessions

- Google Images use common, but with comparatively less emphasis
- Fewer specific filter/feature needs with regard to search behaviors/preferences, despite higher volume of images/image search work and more use cases
- Adaptive Commons search by using categories is widely used to search effectively and avoid the poor search functionality
- Acknowledgment of both the pros and cons of the current Commons set-up and search functionality within
- Majority felt static mock-up provided the best solution (with some assumptions)
- Poor search relevancy and incomplete categorization/tagging awareness exacerbates search

Image search, generally (Usertesting)

What is important for image searching?

- (7) Image source (where it is from) and copyright/reuse/license information
- (2) After selecting an image, seeing similar/relevant images
- (4) size/quality information and filters
- (4) concept chips
- (4) easy download function (the others used right-click to save image file)
- (2) save/add to collection
- Landscape/portrait filter
- Searching with an image (i.e. dragging image into search bar)
- Easy share tool
- Transparent image background filter
- No watermark filter
- Dark mode/skin

https://commons.wikimedia.org/wiki/File:Perspective-Button-Search-icon.png

Image search, generally (Usertesting)

Where and how is the image searching done?

- (10) indicate google images as their go-to
  - (6) it's the only image search used
- (3) actively use concept chips on google images
- (1 each) Unsplash, specifically for non-copyright images, school subscription to Adobe stock photo repository for school presentations, Pinterest for casual aesthetics search, Instagram for casual search

- (10) all use a few words/short phrases in their queries; most believe that longer sentence/question length queries will confuse/overly broaden the result. One felt google images processes additional terms well, and occasionally queries with longer sentences

https://commons.wikimedia.org/wiki/File:Google_2015_logo.svg
https://commons.wikimedia.org/wiki/File:Unsplash_wordmark_logo.svg
https://commons.wikimedia.org/wiki/File:Adobe-Stock-logo.png
https://commons.wikimedia.org/wiki/File:Pinterest_logo.png
https://commons.wikimedia.org/wiki/File:Instagram_logo.svg

Image search, generally (Usertesting)

Other search behaviors

After finding an image of interest:

- (8) right-click to copy or save/download
- Check image size to confirm it will work for the document, presentation, or other media

If nothing is found:

- (3) would adjust/refine query, add/change words to make more specific
- Would search again on Google Images or, if elsewhere such as on Commons, go back to Google Images to search

Current Commons search (Usertesting)

Positives

- (1) has reused Commons images before

Negatives

- (10) all participants had not searched on Commons previously
- (3) specifically mention they don't know what Commons is
- (1) didn't even realize he was on an image search results page

> *Oh, these are the images? Oh I'm so sorry.*
>
> *- Participant*

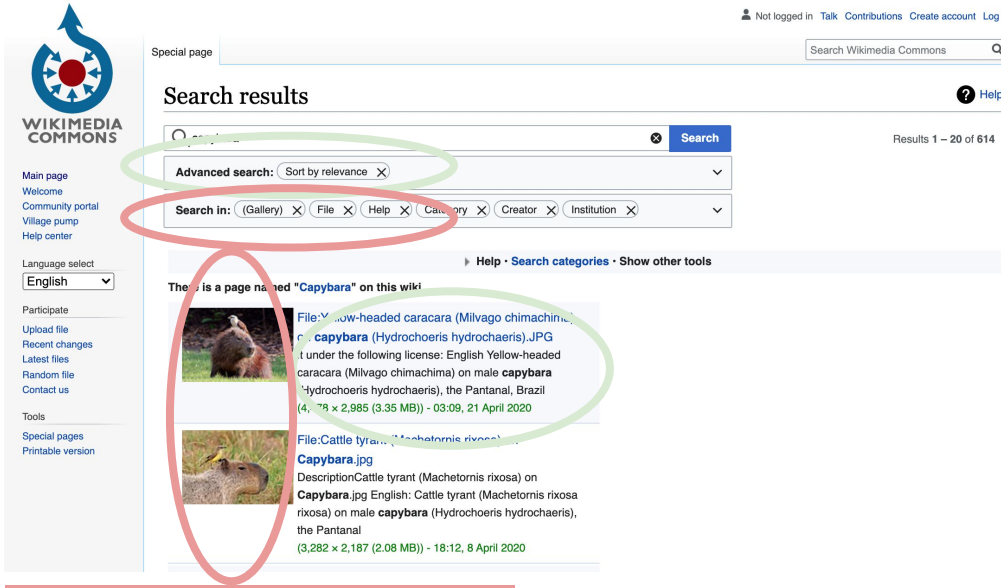
Current Commons search (Usertesting)

Positives

- (2) Lots of details (dimensions, name of file, etc.)
- Advanced search file type dropdown

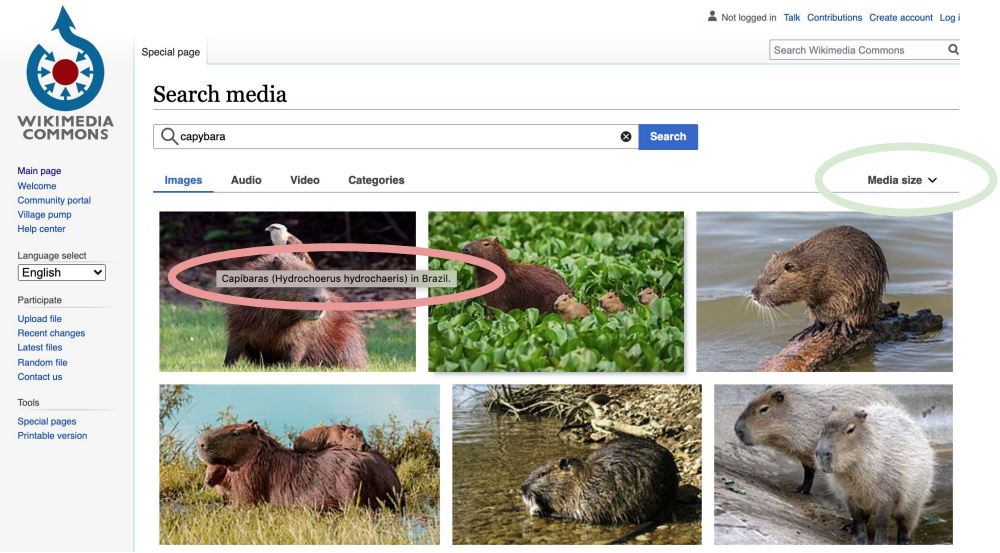
Negatives

- (3) confusion about the ‘search in’ and elements of advanced search
- (2) List format
- (2) Small images “not a very nice display of results”, “difficult to see detail”
- (2) Have to navigate to image page “extra work going to new page”
- Many pages of results instead of infinite scroll
- Too much text
- First images don’t always match query
- Some results have no image previews



> *I don’t trust it.*
>
> *- Participant*

Prototype search (Usertesting)



> *Looks more like google image search.*
>
> *- Participant*

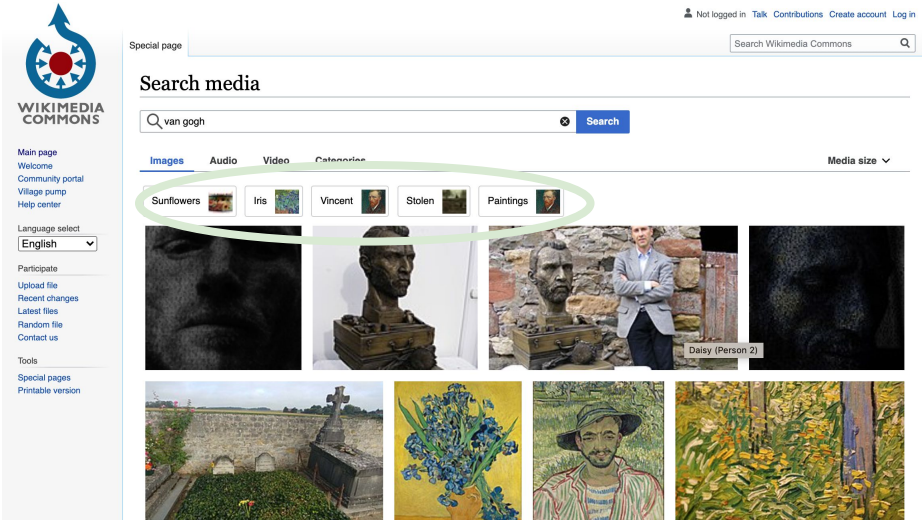
Positives

- (9) looks better
- (2) Clearer, larger images
- Size filter
- infinite scroll

Negatives

- (3) Have to navigate to image page (and back)
  - Would like if it at least opened image page in new tab
- Would like to be able to see resolution and other details on hover
- No advanced search like current commons search, option to choose file format
- Search suggestions are not helpful
- Images still not as well matched to query as google images' are

Concept chips (Usertesting)



> *Think it's useful to have; provide[s] direction when doing open-ended research.*
>
> *- Participant*

(9) Users believe they are meant to narrow/refine search results.

Positives

- (3) noticed the chips; “they're nice, provide links to associated [topics]”
- (10) would be good to include them as feature
  - Would like them to be constantly present on search, like on Google Images where clicked chips are ‘active’ and those not clicked aren’t
- (5) helpful to focus search “recommendations for keywords to be added to my search” “you can filter your search”

Negatives

- (7) did not notice them
- Didn’t think the chip would just add the word(s) to the query
- Typically wouldn’t use

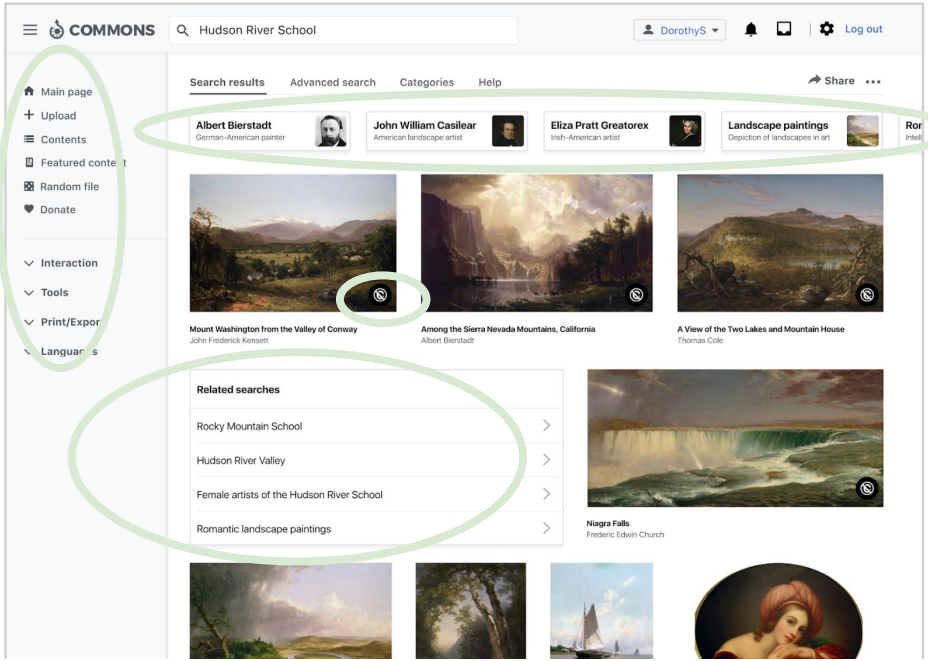
Static search mock-up (Usertesting)

Positives

- (9 of 9, image didn't load for 1 user) clean interface pretty.
- (5) concept chips look better/have more information
- (4) like advanced search and tools on left
- (2) seems like you could maybe see info on the images without leaving the page
- (2) like related searches box
- like the copyright icon

Negatives

- n/a

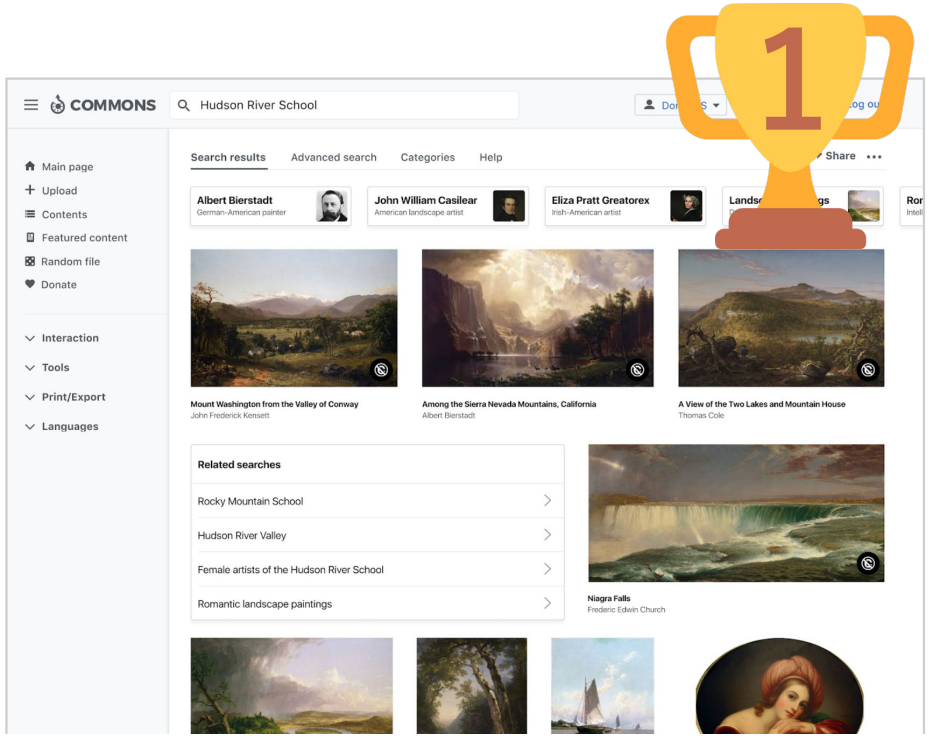


> *Sleek. Clean. Modern.*
>
> *- Participant*

Preferred search interface (Usertesting)

- (8 of 9) liked design of the static mock most
  - (2) would want to be able to know what is in advanced search / search audio/video/image in mock option
  - (2) want to be able to enlarge image on the same page, generally have access to as much information as possible w/o leaving search page
  - Related; users did not want to navigate from search to image and back; the assumption is that the static mock would behave similarly in that sense to Google Images

- (7) current commons search the worst “the first one was a total mess”
  - Current commons search list view is limiting, and search results not relevant



https://commons.wikimedia.org/wiki/File:1st_Place_Trophy.png

Image search, generally (Moderated)

What is important for image searching?

- (2) reusable/open license filter
- Color filter
- Date/recency filter
- Image size/resolution filter
- image quality information/filter

How is the image searching done?

- (5) typically write queries as short phrases
- Initiatives and tasks undertaken: Wiki Loves Monuments/Earth, GLAM, Commons admin (1), mass uploads of all media (3), upload of personally captured media (3), image categorization/adding metadata (3), monitoring image use on articles, move open license images from Flickr
- Tools used: Upload Wizard (3), Commonist, Vicuna, Pattypan, Catalot, FlickrtoCommons
    - Would like to be able to select templates (e.g. photograph, artwork, etc.) as options on Upload Wizard, instead of just the default information template

https://commons.wikimedia.org/wiki/File:Perspective-Button-Search-icon.png

Image search, generally (Moderated)

Where is the image searching done?

- (3) indicate Google Images as their go to (with free license filter)
- (3) Commons search via categories, subcategories
  - (1) uses Commons as preferred image search
- (2) Flickr
- No longer uses google images because the access to the image itself is now limited/difficult to access
- Uses Creative Commons
- Uses Unsplash, Pixabay

https://commons.wikimedia.org/wiki/File:Google_2015_logo.svg
https://commons.wikimedia.org/wiki/File:Flickr_logo.png
https://commons.wikimedia.org/wiki/File:Unsplash_wordmark_logo.svg
https://commons.wikimedia.org/wiki/File:CreativeCommons_logo_trademark.svg
https://commons.wikimedia.org/wiki/File:Pixabay-logo.svg

Image search, generally (Moderated)

Other search behaviors

After finding an image of interest:

- (5) would review license information if needed
- (2) would download original file/best resolution version
- Would right-click and copy
- Would check if the image is available in other languages

If nothing is found:

- (5) would adjust query

Current Commons search (Moderated)

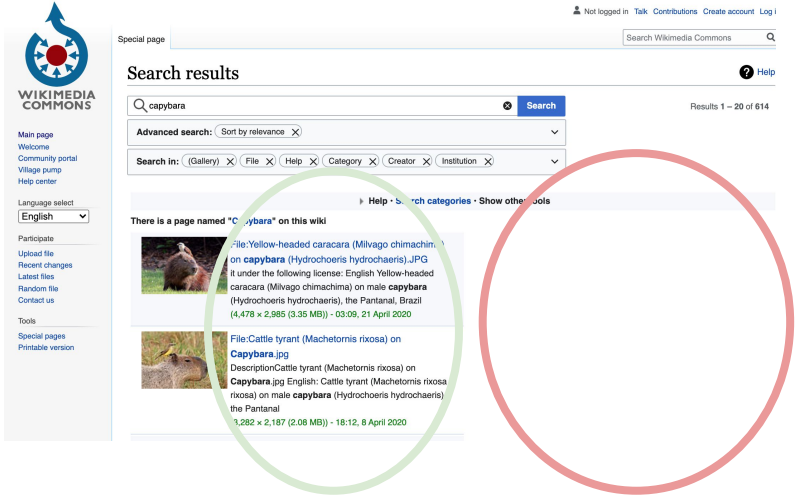
(5) All users from this cohort had strong inclinations to search by category instead of using the search bar. Awareness is high that unless one knows the exact file name, search result relevancy is low.

Positives

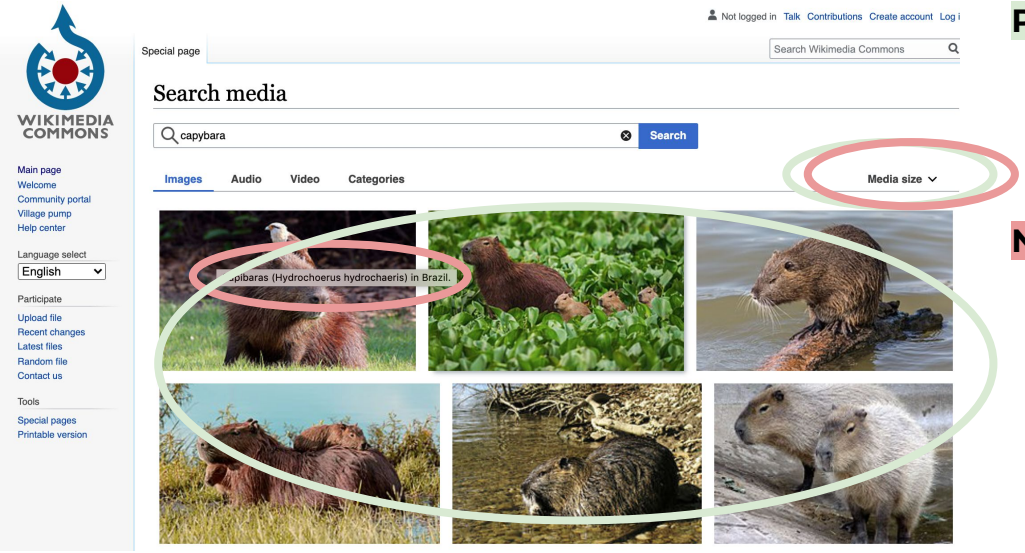
- Description and image information is important

Negatives

- Too much white space



Prototype search (Moderated)



Positives

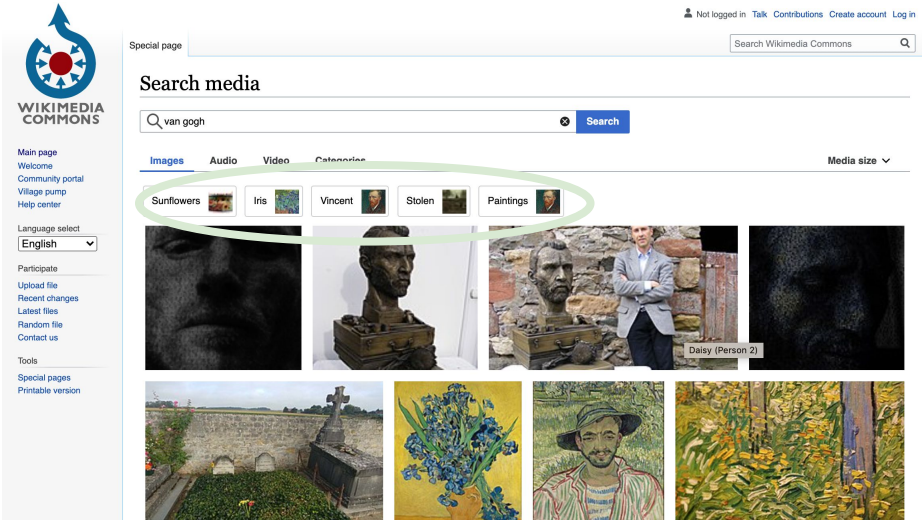
- (4) more images, good visual display
- (4) media size filter
- (1) infinite scroll

Negatives

- (4) more information on hover, title/description, resolution, etc.
- (2) unclear what the sizes represent on the media size filter dropdown
- Player/scrubber is too short on audio tab results
- On video tab results, want length of video and resolution information
- Category tab; would want to see a category tree
- Prefer intermediate step when clicking on images (vs, click-through to image page)
- See related images after clicking on an image

> *When searching for media, better to show media.*
>
> *- Participant*

Concept chips (Moderated)



(5) felt they were meant to filter/make the results more precise

Positives

- (5) noticed them
- (3) had positive reactions

Negatives

- (2) prefer that they behave like swatches that could be tagged on and off, like the Google Images chips
- Felt results were already not always accurate, and that adding words may not help

> *Good improvement.*
>
> *- Participant*

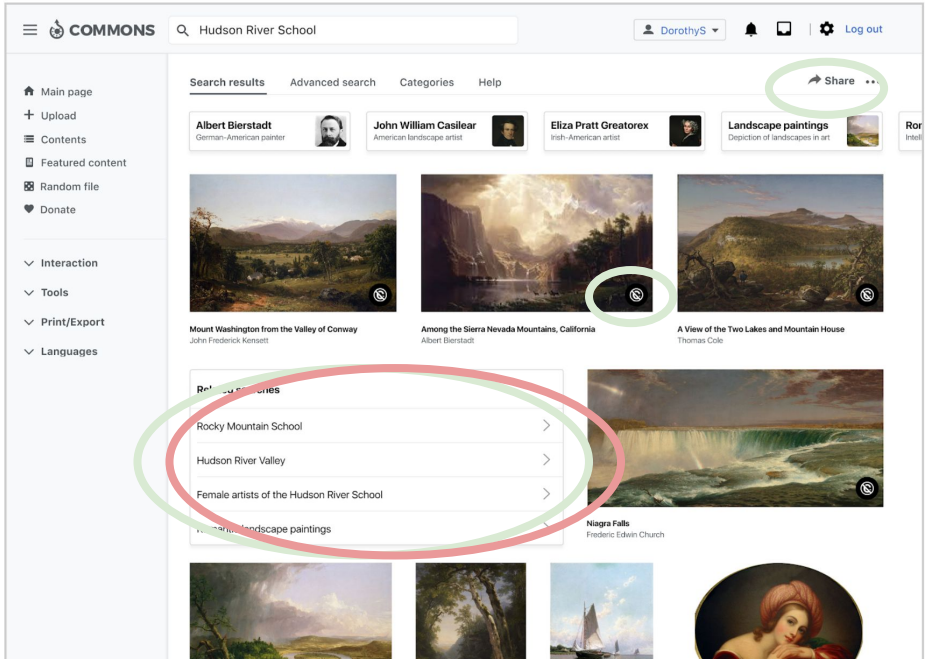
Static search mock-up (Moderated)

Positives

- (4) had positive responses to visuals
- (3) liked the license icon
- (3) liked related searches
- Likes share capability
- Would want categories tab to display a list of categories to which image belongs (corresponding to each image represented as a thumbnail)

Negatives

- Didn't like the location of related searches

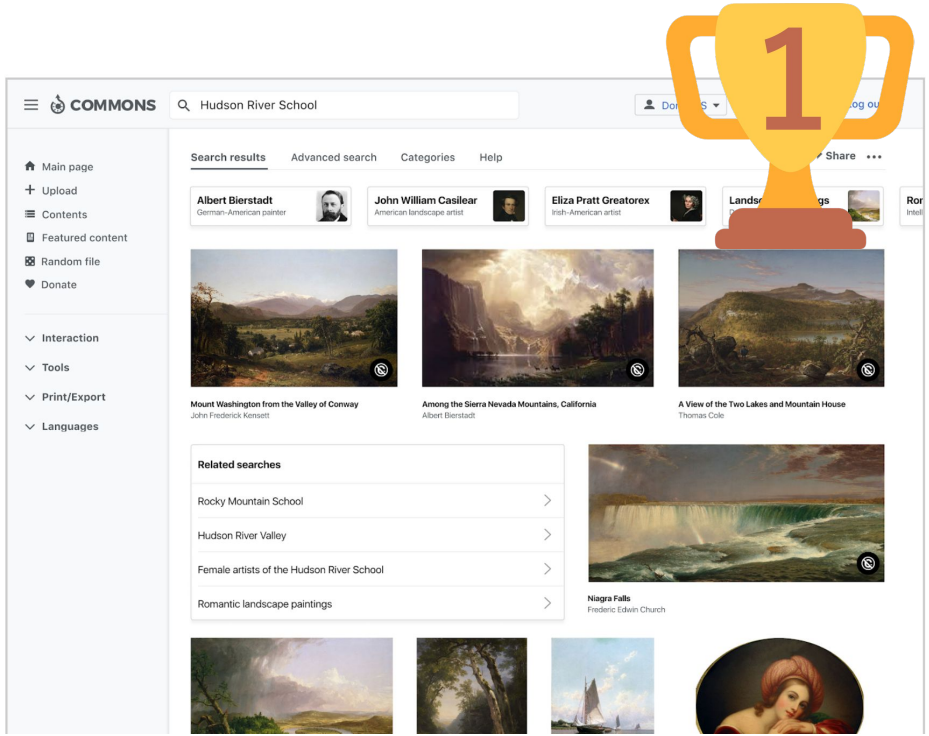


Preferred search interface (Moderated)

(4 of 5) had clear preference for the static search mock.

One user felt that the prototype search is not a good middle ground between current and the static mock, and that the best would be a strategic merge of the current Commons search and the static mock best qualities.

Another user mentioned wants to make sure image searching via categories will remain easy to do if a new search is implemented.



https://commons.wikimedia.org/wiki/File:1st_Place_Trophy.png

Recommendations

- **Search relevancy**;
  - A refrain from these sessions was that image searches did not return results matching the query well. The best solution to improving search on Commons needs to be a **holistic one, and in this case, improving the UI without improving the search functionality itself is, perhaps, like putting new paint on a crumbling house.** In conjunction with improving the search UI, other **internal factors like working on structured data and improving the algorithm behind Commons search** combined with other factors such as **additional pushes for media upload type edit-a-thons** (both of photos taken by individuals and 'mass import' of open source media type events), **photo competitions, mass uploads** from members of specific communities (e.g. academia, museums, research institutions, etc.) can help solve this problem in a two-pronged way.
  - Related: for experienced Commons users, not only was search relevancy an issue with search (low image availability or quality due to algorithm omission/deprioritization), but that images were further 'hidden' due to poor categorization and tagging. When this is the case, the grunt work of completing file metadata, caption/description, can prevent low/irrelevant search results even with the category search workaround.

> *First images don't always match query*
>
> *- Participant*

> *Images still not well matched to query*
>
> *- Participant*

Recommendations

- A few users mentioned the idea of a site identity, and that the current search matches the ‘wiki look’ but that it is outdated. Finding the midpoint or merging the high-value features of each possibility and retaining brand identity is important moving forward.
- Experienced users requested that the category search workaround be left intact, even if a new search were to be implemented. Additionally, providing categories for each image search result and upon further exploration, being able to view category hierarchies, would provide value.
- Image quality tags/badges, similar to article designations such as ‘Featured’ would be helpful.
- Explore the idea of toggle on/off concept chips that are constantly present when a query is entered.
- After clicking an image result, additional details and related images on an intermediary page/pop-up should be shown to reduce the need for navigating to/from image pages.
- Provide additional information when hovering over images.
- Show length of video and resolution information on video results.
- Show longer players for audio so the scrubber is functional.
- Consider the inclusion of ‘nice-to-have’ filters/features not currently available.

Next steps

- Improving search on Commons should continue in the form of continued iterative testing of prototypes, perhaps a built-out version of the static mock from this round.
- Future endeavors for this team will include improving search on not just Commons, but all wiki projects (Wikipedias in particular).

Research Approach

- Objectives
- Hypotheses/Questions
  - Answers to same
- Methodology
- Participants

WIKIMEDIA
FOUNDATION

Objectives

Objectives

The goals of this research project are three-fold:

1) Understand a range of user behaviors related to image searching on the web
2) Evaluate users' understanding of and experiences with the 3 search environments
3) Synthesize findings from user feedback into recommendations that will steer the larger goal of improving search on Commons

Hypotheses/Questions

Hypotheses & related questions

HYPOTHESIS 1: users overall will find a more image-based and less text-heavy version of Commons search easier and more useful.

- Are there any benefits of one approach vs the other that have not been highlighted?
- What is the best way to meld the best of both worlds?

HYPOTHESIS 2: a search experience driven by structured data (specifically 'depicts' statements) will lead to search results that users find more helpful than traditional search results on Commons.

- Will this be the case for all users? Will differing needs and expectations need to be met?

HYPOTHESIS 3: different users will have different expectations; some will want to keep the status quo, others will want change and modernization.

- Other than keeping a legacy version of search, what are the other ways we can address a wide variety of use cases within the new search framework?

Hypotheses/Questions - some answers

Hypotheses & related questions

HYPOTHESIS 1: users overall will find a more image-based and less text-heavy version of Commons search easier and more useful.

- Are there any benefits of one approach vs the other that have not been highlighted?
- What is the best way to meld the best of both worlds?

While the reaction to the image-based search experiences was very positive, there are still elements of the current Commons search (and greater experience) that users would like to retain. Further iterative testing of search prototypes should reflect and retain these elements as much as possible.

Hypotheses/Questions - some answers

Hypotheses & related questions

HYPOTHESIS 2: a search experience driven by structured data (specifically 'depicts' statements) will lead to search results that users find more helpful than traditional search results on Commons.

- Will this be the case for all users? Will differing needs and expectations need to be met?

To better ‘prove’ the efficacy of structured data in the search context, an A/B test or, if available, reviewing any instrumentation data from analytics can better illustrate impact/effectiveness.

However, structured data is generally a boon to finding media to display and finding related media. In addition to evaluating effectiveness and fine tuning, we also need to address the search algorithm itself and encourage efforts to make the data/content itself more robust both in terms of MORE content, period, and its available metadata when added to Commons and the wikis. It can’t help if it can’t be found.

Hypotheses/Questions - some answers

Hypotheses & related questions

HYPOTHESIS 3: different users will have different expectations; some will want to keep the status quo, others will want change and modernization.

- Other than keeping a legacy version of search, what are the other ways we can address a wide variety of use cases within the new search framework?

Of the casual searcher Usertesting cohort, the majority, as expected, reacted very well to the more modern, clean, image-based search interface.

Among the seasoned Commons veterans, there was a surprising amount of openness to the potential of a search interface overhaul. Areas where the current search fail and optimism and excitement about the prototype/mock searches (of course, with some assumptions about functionality where applicable) were freely discussed.

With the feedback provided, there are viable paths to moving forward on modernizing search while still allowing for searching via category independently, or within the context of the search interface.

Methodology

Methods & approach

- Research protocol developed to address users' image search use cases, image search behavior, and the 3 search environments (current Commons search, prototype search, search static mock)
- Both unmoderated sessions and moderated sessions were utilized to quickly capture general, casual image searchers (Usertesting.com cohort) and dedicated Commons users
- Progress updates and write-ups delivered to SDAW team on an ongoing basis

Research sessions

- 10 unmoderated remote sessions via Usertesting.com
    - Each session followed the same format; 1) pre-test screener survey to ensure wide range of demographics 2) overview of image search behaviors 3) review of Commons, prototype and static mock 4) post-test questionnaire
- 5 moderated remote sessions using Google Meet
    - Each session followed the same format; 2-4) of the above

Participants (Usertesting.com)

Participant recruitment

- Tests were set up to promote geographic and gender diversity among participants from the Usertesting panel

Data collection & analysis

- Approximately 4 hours of video recordings from 10 sessions
- Each finding is recorded and tallied to establish patterns in behaviors, responses to the 3 search environments, and related preferences

Participant Overview

- Per product manager request, these users primarily image search in educational contexts (high school and college students)
- 6 women, 4 men
- Countries represented:
    - (4) USA
    - (2) New Zealand
    - (2) Mexico
    - (1) Australia
    - (1) Philippines
- Ages represented: 18, *(2)* 19, 20, *(3)* 21, 22, 23, 36

Participants (Moderated)

Participant recruitment

- Legacy contact spreadsheets from past projects with Commons users, GLAM representatives, external re-users, etc., were leveraged to recruit participants

Data collection & analysis

- Approximately 5 hours of video recordings from 5 sessions
- Each finding is recorded and tallied to establish patterns in behaviors, responses to the 3 search environments, and related preferences

Participant Overview

- 2 women, 3 men
- 3 GLAM representatives, 2 heavy Commons users

Discussion

Questions, Comments, Feedback

Thanks

Research participants, for your feedback

Ramsey Isler, Matthew Williams and Carly Bogen, for your input and collaboration

WIKIMEDIA
FOUNDATION

Fin

Product Design Strategy

Direct questions & comments to:
dchen@wikimedia.org

WIKIMEDIA
FOUNDATION