



HATHI™
TRUST

PROJECT

HathiTrust Personas

The Making of the HathiTrust Personas

PROJECT
COLLABORATORS
/AUTHOR(S)

By Sonali Mishra (project research, personas, and report),
under advisement of the HathiTrust User Experience
Advisory Group: Suzanne Chapman (chair), Darcy Duke,
Jenny Emanuel, Michael Lascarides, Felicia Poe

SHORT
DESCRIPTION

In 2011, the HathiTrust UX Advisory Group developed a set of personas to represent users of the HathiTrust Digital Library. The resulting seven personas draw from past usability evaluations, user research, user-generated comments and feedback, web analytics, observations of use collected by the UX Advisory Group and their colleagues, and other sources. These personas aim to help staff working on HathiTrust (developers, policy makers, user experience designers and researchers, reference and instruction librarians, etc.) learn more about HathiTrust users, discover how we can better suit their needs, and identify areas in which to do more in-depth research.

REPORT INFO

Report Date: 11/11/11
Revision: 1.0
Contact: Suzanne Chapman (suzchap@umich.edu)
Available online: <http://hathitrust.org/personas>

Table of Contents

<i>Table of Contents</i>	2
<i>Introduction</i>	2
<i>Research</i>	2
Analytics	2
Anecdotes	3
Articles	3
Blogs/Comments/Forums	3
Error Typology	4
Research from past projects	4
Unsolicited User Feedback	5
<i>Creating the Personas</i>	5
<i>Discussion</i>	6
Limitations	6
Areas for Future Work	6

Introduction

In 2011, the HathiTrust UX Advisory Group developed a set of personas to represent users of the HathiTrust Digital Library. The resulting seven personas draw from past usability evaluations, user research, user-generated comments and feedback, web analytics, observations of use collected by the UX Advisory Group and their colleagues, and other sources. These personas aim to help staff working on HathiTrust (developers, policy makers, user experience designers and researchers, reference and instruction librarians, etc.) learn more about HathiTrust users, discover how we can better suit their needs, and identify areas in which to do more in-depth research.

Research

Information about users was culled from web analytics, blogs, user feedback, staff anecdotes, and notes from previous research projects. Although not everything we found made its way directly into the personas, all the information we found informed our understanding of our users.

Analytics

Web analytics furnished us with knowledge about where our users were located, what kinds of topics they searched for, and what parts of the system garnered the heaviest use. For example, we discovered that many of our visitors were based in Austin, TX (possibly because of the university there), while an analysis of search terms conducted in 2010 revealed several common categories of queries, including

genealogical queries, handicraft-related queries, and queries about boating or ship construction.

- Search term analysis conducted by Sonali Mishra in Winter 2010 on search queries from 1/1/10-1/31/10. Explored catalog search queries to find common search topics.
- Google analytics reports from Jan 1- Mar 31 2011. A variety of reports were mined for information like top referring sites, top locations from which users came to the site, and top languages in which the site was viewed.

Anecdotes

Real life stories are not available for public viewing. These stories were gathered from library staff by the UX Advisory Group to inform us about how librarians interacted with the system, personally and professionally, and about how they helped patrons with its use. For example, one librarian recounted his/her experience using full-text search as a surrogate index for unindexed documents, while another told us they had used collection builder as a way to identify works for deaccessioning during the consolidation of two libraries. Another librarian relayed their experience helping a patron view some marginalia that he had spotted in one of the scanned pages of HathiTrust, which would not have been available to him from OCR alone. We also heard a nice anecdote about a librarian's personal use of the site – they had found an old cookbook on the site and had it custom printed on the University of Michigan's Espresso book machine as a gift for a friend.

Articles

- Rich, Motoko. "Google hopes to open a trove of little-seen books." New York Times, 1/4/2009
http://www.nytimes.com/2009/01/05/technology/internet/05google.html?_r=1 Describes lexicographical research in Google Books

Blogs/Comments/Forums

Blog posts written by users from various walks of life were a useful source for learning about who our users are, what they came to HathiTrust to do, what they liked or disliked, and sometimes whether they were academic users or not. These blog posts also occasionally gave us some insight into why users chose HathiTrust over other sources. For example, one post in Language Log, a linguistics blog with faculty and graduate contributors, described a linguist's experience of researching the history of a particular joke. He found an early occurrence of it in print by using HathiTrust's full-text search, and said HathiTrust's metadata about the volume was more accurate than Google Books's. Another blog post illustrated a non-

academic use of HathiTrust – one Armenian food blogger posted about an ad for a “portable Armenian kitchen” found in a 19th century ladies’ magazine in HathiTrust, asking her readers if they had ever seen anything like it. A selection of posts used to inform the personas:

- Language log, 9/16/2008 <http://languagelog ldc.upenn.edu/nll/?p=602>
Describes using HathiTrust and Google Books to research the history of a joke; praises Hathi’s metadata
- <http://tinyurl.com/7guwvr8> Classics librarian describing use of collection builder to make some classics themed collections.
- <http://minneapolisparkhistory.com/2011/03/10/minneapolis-park-history-resources-hathitrust/> Nice introduction to HT for people researching Minneapolis history, particularly parks.
- <http://typophile.com/node/79276> Used HathiTrust item to ask for help identifying an old typeface
- <http://m3jfm.net/?p=1223> Discovery of Public Domain travel guides
- <http://www.thearmeniankitchen.com/2010/11/portable-armenian-kitchen-19th-century.html> Discovery of an Armenian cookbook in HT
- <http://www.earlyamericancrime.com/convict-transportation/convict-voyages/convict-attitudes> HathiTrust item citation.
- <http://ichthyscybernetics.blogspot.com/2010/10/john-wesley-and-extraterrestrial-life.html> Uses material in HathiTrust in a discussion of John Wesley’s take on extraterrestrial life.
- <http://searchinfor kinfolk.blogspot.com/2011/04/hathitrust-digital-library.html> Description of genealogy search

Error Typology

The error typology is a list of types of quality errors encountered on the site. It details problems with page scans and OCR. It was developed by HathiTrust Copyright Review project staff.

Research from past projects

Research from past projects was mined for information about who our users were, where they came from, and what they intended to do on the site, as well as general background information.

- WCL Survey: This survey was created to gauge users’ feelings about the worldcat local catalog prototype. As part of this survey, participants were asked to identify themselves as graduate, undergraduate, library staff, faculty, and other roles, and were asked to state their reasons for coming to the site and what they planned to do with their findings.

- University of Michigan MBooks Interviews (HathiTrust predecessor): notes from interviews done in preparation for the MBooks project.
- University of Michigan Mirlyn Library Catalog: Comments made by study participants in a study on Mirlyn's default exclusion of limited view items.

Unsolicited User Feedback

User feedback is not available for public viewing. This feedback includes comments from users about what they like and dislike about HathiTrust. These emails frequently include contextual information about what the user was trying to research or accomplish.

Creating the Personas

The data culled from these sources was analyzed and compiled into personas. The first step was to segment the users into groups with different goals. In segmenting our users, we first divided users up as academic or non-academic, then examined sources like responses to a WorldCat Local survey conducted earlier in the year, a search term classification conducted in winter of 2010, and anecdotes about academic and non-academic users to isolate different user types and goals. The final segmentation was as follows:

- Non-academic users
 - hobbyists
 - genealogists
- Academic users
 - faculty
 - graduate students
 - undergraduate students

At least one persona was made to represent each of these groups. Based on the data we gathered, we developed each persona's interests, goals for using HathiTrust, likes and dislikes about the system, and challenges in using HathiTrust. In some cases our data allowed us to assign these attributes to specific user groups with some certainty; in other cases, we hypothesized based on what we had learned. Finally, we added additional details about their personal backgrounds to complete their stories.

Discussion

Limitations

Our method of using past research, a low-cost approach, to create the set of personas came with obvious limitations.

With the exception of the web analytics data, most of what we gathered was primarily from the most vocal users -- both the positive and negative extremes -- or from a self-selected sample of users asking for help at the reference desk. These users may or may not be representative of the entire user body. We also had more data about academic users than about non-academic users, simply because academic users are more accessible.

The web analytics data we gathered provided a nice balance to the rest of the sources and was useful for providing aggregate figures and search terms. However, it did not tell us much about individual user goals and satisfaction -- the reasons for the behavior. For example, if we were to learn that the average amount of time spent on full-text search was higher than on the rest of the site, we would not know if it was because users were delving deep into full-text search, or because they were encountering usability issues that took time to work around. Although we have gained some insight into where users are satisfied or frustrated by anecdotes, blogs, and user feedback, we cannot cleanly map these reasons onto data from web analytics, and we cannot be sure if the insights we gained are representative of the entire user body.

Since much of the data we gathered was from anonymous sources, we also cannot always know which findings belong to which user groups. This matters because user groups may need different routes to accomplish their goals. For example, if we have a cluster of users in Austin, TX, it is certainly plausible that they are all university affiliates; however, it is equally possible that they are actually unaffiliated with the university. If our Austin users are largely affiliated with the university, they are more likely to have access to limited view items through their own university library. However, if they are not affiliated with a university, they may not have direct access to the print materials. In creating our personas we made our best guess in assigning findings to particular user groups; however, we cannot know for certain if we were correct.

Areas for Future Work

A number of possibilities for future investigation come out of this project, including:

- Validation and enrichment of findings: further investigation can help us evaluate and deepen our findings on user goals, user behavior patterns, and user segmentation. Are the goals we found through our research representative of other users? Are the categories of users we used valid? And can we deepen our understanding of which users use which parts of the system, and how they use it?
- International users: although a preliminary persona was created for an international user, we soon discovered that we knew too little about these users to flesh him out properly. Further research may help us determine if HathiTrust is meeting the needs of a global audience.
- User satisfaction and retention: we have very little data on user retention due to limitations on the information that web analytics can capture. Further research could explore what factors correlate to user retention, or even whether user retention should be a primary goal for HathiTrust.
- Competition: although HathiTrust is a non-profit initiative, it may still be beneficial for HathiTrust to analyze why some users choose competing services, like Google Books, over HathiTrust. Although our goal is not to grab market share from competing services, our goal is to meet the needs of a public audience, which may be either academic or non-academic. Comparative research may help us discover how we may better meet the needs of this audience.