Google Books' Coverage of Hawai'i and Pacific Books

Andrew Weiss
The California State University, Northridge
Oviatt Library
18111 Nordhoff St.
Northridge, CA 91330
andrew.weiss@csun.edu

Ryan James
The University of Hawai'i at Manoa
University of Hawai'i at Manoa Library
2550 McCarthy Mall
Honolulu, HI 96822
rsjames@hawaii.edu

ABSTRACT
This poster reports on a recent quantitative study of *Google Books'* coverage of Hawaiian and Pacific books. A total of 1,500 books were randomly selected from the University of Hawai'i at Mānoa's Hawaiian, Pacific, and general stacks collections. Their level of access was then determined in *Google Books* by observing whether the books had a metadata record, were full-text searchable, and whether they were available in snippet, preview, or full-text views. Results show that *Google Books* has a sizable number of metadata records for Hawaiian and Pacific books, but has only a limited number available for full-text searching. In contrast, a larger number of books from the general stacks were available for full-text searching.

Keywords
Google Books, metadata, massive digital libraries, information access, organization of information

INTRODUCTION
The investigators have taken Euro-centric concerns of diversity and applied them to historically under-represented groups in the United States (Bearman) (Jeanneney). Going back to the original question of whether *Google Books* is a universal library or even an American universal library as Jeanneney in his text *Google and the Myth of Universal knowledge: A view from Europe* claims, the question the researchers ask is whether *Google Books* is a "*mainland American* universal library"? The researchers believe it is worthwhile determining how closely *Google Books* reaches its stated goal of universality.

The University of Hawai'i at Mānoa Libraries have a long history of prioritizing materials relating to Hawaiian and Pacific cultures. UH-Manoa librarians have spent decades developing these collections. The result of their effort has been a collection of over 150,000 items in the Hawaiian collection, and a Pacific collection of over 130,000. This is one of the largest and most comprehensive collections relating to these subjects in the world (Department History, Special Collections). It also represents an important starting point for assessing Google Books' universality and comprehensiveness.

METHODOLOGY
A random sample of 500 books was taken from three collections each, the Hawaiian Collection, the Pacific Collection, and the general stacks collection, for a total of 1,500 books.

Selected item records were then searched in *Google Books*. The primary metadata fields used for searching were title, author, and publication date. Metadata related to page number, publisher, edition and reprint were also collected for identifying book records in *Google Books.*

Matching item records found in *Google Books* were then evaluated for their level of access. The categories used were 'Record,' 'Snippet,' 'Preview,' and 'Full'. The 'Record' category signifies that *Google Books* only provides access to a metadata record for the selected item. 'Snippet' provides the ability to search the text of the item but displays only a tiny portion of the text. 'Preview' also allows for full text searching but allows for only a larger, though still limited, number of pages to be viewed. 'Full' offers full-text searching and access to the entire item for unrestricted viewing.

Selected items records were recorded at their highest level of access, and only categorized once. Additionally, selected item records that could not be found in *Google Books* were recorded in the "no

record" category. There are no inter-coder statistics available because all coding and categorizing of the data were done by one of the investigators.

RESULTS

- Of the 500 randomly sampled Hawaiian collection books, 131 had no record in Google Books, 317 had a metadata record only, 35 had a snippet view available, 11 were available for preview, and 6 could be fully viewed.

- For the Pacific collection books, of the 500 randomly sampled books 116 had no record, 261 had a metadata record only, 98 had a snippet view, 17 were available for preview, and 8 could be fully viewed.

- For the general stacks collection, out of 500 randomly sampled books 40 had no record, 153 had a metadata record only, 229 had a snippet view, 66 were available for preview, and 12 could be fully viewed.

- Additionally, 6 Hawaiian, 6 Pacific collection and 29 general stacks collection books were available for purchase through Google Books.

- Of the 21 total books in the subset of sampled Hawaiian collection books that likely fall within the public domain, 4 of these had no record, 11 had only a metadata record, none (0) had a snippet view and 6 had a full view (29% of titles).

- Of the 24 total books in the subset of sampled Pacific collection books that likely fall within the public domain, 4 of these had no record, 11 had only a metadata record, 3 had a snippet view, and 6 had a full view (25% of titles).

- Of the 18 books in the subset of sampled general stacks collection books that likely fall within the public domain, 3 of these had no record, 3 had only a metadata record, 1 had a snippet view, and 11 had a full view (61% of titles).

- None of the collections had any books available for preview from this subset of the sample.

DISCUSSION

In the sample of books taken, the general trend shows that there are a greater number of items that either have no record in *Google Books* or have only metadata records, suggesting that these books have not been digitized. Furthermore, a larger number of books with snippet views exist for a general collection versus the Hawaii and Pacific collections. In fact, there exists a 2.34 to 1 ratio between the books with snippet views in the General collection and those in the Pacific collection; this increases to 6.54 to 1 between the general stacks and the Hawaiian collection. Books with a preview are also tilted more toward the general stacks collection as well, with a 3.88 to 1 ratio between general stacks books and the Pacific collection and a 6.00 to 1 ratio between general stacks and the Hawaiian collection. Finally, the number of books with full views becomes a little more evenly represented with the ratios flattening to 1.5 to 1 for stacks and Pacific books and 2.0 to 1 for stacks and Hawaiian books.

As far as works in the public domain are concerned, one might expect more evenly matched numbers. This is not the case, however. Despite the possibility of open access and the lack of copyright restrictions, the ratio of books available for full view in the general stacks is still nearly twice that of both the Hawaiian and Pacific books (1.83 to 1). Indeed, when looking at the number of books with metadata records only and those with no records at all, it is obvious that more than half of the books have not been digitized.

CONCLUSION

The results of this research suggest that *Google Books'* approach to the mass digitization of books does not provide the adequate coverage of diversity from multiple cultural perspectives – the Hawaii and Pacific in particular – and raises the issue of further marginalizing underserved groups.

The results point toward a large majority of the books in the Hawaiian and Pacific collections not being digitized at all. Metadata records exist for many of

these items, but with no views, and in some cases no records of any kind available, one must assume that digitization has not taken place.

After looking at these results, we can see that despite a collection of over 15 million books (just 15% of the estimated 100 million books in Western culture), (Jeanneney) when looking at the nature and quality of the *Google Books* corpus, there is much to be desired compared to the Hawaiian and Pacific collections of a library specializing in these subjects.

REFERENCES

Bearman, D. (2006). Jean-Noël Jeanneney's Critique of Google: Private Sector Book Digitization and Digital Library Policy. *D-Lib Magazine*, 12 (12). Retrieved June 8, 2012 from http://www.dlib.org/dlib/december06/bearman/12bearman.html

Department History, Special Collections. University of Hawaii at Manoa Libraries. Unpublished.

Jeanneney, J. (2006). *Google and the Myth of Universal knowledge: A view from Europe*. Translated by Teresa Lavendar Fagan. Chicago: University of Chicago Press.