

Analysis of Transaction Logs for Insights into Use of Life Oral Histories

Michael G. Christel

Entertainment Technology Center
Carnegie Mellon University
Pittsburgh, PA 15213
1-412-268-7799

christel@cs.cmu.edu

Bryan S. Maher

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
1-412-268-8970

bsm@cs.cmu.edu

Huan Li

School of CS and Engineering
BeiHang University
Beijing, China

huanli@andrew.cmu.edu

ABSTRACT

A digital video library of over 900 hours of video and 18000 stories from The HistoryMakers was used by 214 students, faculty, librarians, and life-long learners interacting with a system providing multiple search and viewing capabilities over a trial period of several months. User demographics and actions were logged, providing metrics on how the system was used. This poster overviews a few highlights from these transaction logs of the Informedia digital video library system for life oral histories.

Categories and Subject Descriptors

H.3.7: Digital Libraries – *system issues, user issues*; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems –*video*

General Terms

Design, Human Factors

OVERVIEW

The Informedia research group at Carnegie Mellon University has worked with an oral history archive, The HistoryMakers, to apply automated techniques and generate time-aligned metadata for use in accessing video narratives [1]. This poster reports on transaction logs of the Informedia Digital Video Library (IDVL) system collected in 2008 across a number of institutions. 214 participants (129 female, 85 male) generated logs, with 176 using IDVL only once; stated occupations were 130 students, 26 not provided, 14 administrators/managers, 7 librarians, 7 professors, and 30 varied occupations repeated at most twice. All but 2 participants had at least a high school education, with 138 having at least some college education and 19 having some graduate school education. They were somewhat experienced web searchers but less experienced video searchers.

288 sessions were logged, representing 234 hours of use: 214 first-time sessions and 74 sessions from a returning user. First-time sessions were characterized by more text queries with less precision (larger answer sizes). Users then spent the majority of their time selecting video stories from the answer sets and playing them. Return users still spent a majority of their time playing video, but experimented more with advanced search options and the various views into video sets. 83% of self-described students never moved beyond the default view, using no optional views. Investigative users accessed timeline and common text phrase

This material is based on work supported by the National Science Foundation under grant IIS-0705491.

Copyright is held by the author/owner(s).
JCDL'09, June 15–19, 2009, Austin, Texas, USA.
ACM 978-1-60558-322-8/09/06.

views the most, but still less than once on average per session. Overall, the participants found their experiences with the system satisfying, historically meaningful, inspiring, entertaining, and accurate based on scores at or near 4 out of 5 on subjective scales presented in a concluding questionnaire.

For 1138 text queries returning results, two human judges assigned the queries into the Panofsky/Shatford framework of specific-generic-abstract [2]. Each query was classified into one of the categories listed in Table 1. Some error was introduced in forcing compound queries (e.g., “Oliver Brown vs. Board of Education”, a specific person, organization, and event) into a single category, but the exercise was conducted to get a general view into how the HistoryMakers was being queried. The two judges classified queries independently, and their classifications agreed 96% of the time. The resulting classification of the first judge is shown in Table 1. Note how few queries were of the abstract type, queries like “overcoming” and “truth beauty.”

Table 1. Human classification of text queries producing results in IDVL for the HistoryMakers Digital Library, given as percentages (of the 1138 total queries classified).

Specific				Generic		Abstract
Location	Person	Event	Other	Event	Other	
10	29	9	6	2	42	2

301 text queries were issued returning zero results, with 130 on content not in the corpus, e.g., “LeBron James.” Another 71 were valid queries that returned zero results because of two known shortcomings with an early IDVL search service, both addressed in the subsequent move to the Lemur Toolkit and Indri search engine. Another 65 queries returned no results because the syntax to directly type in a search against only a portion of the corpus, e.g., to just PoliticalMakers, was mistyped. The advanced search link led to other means of searching, but the shortcuts and punctuation were nonstandard with the major web search engines and led to confusion. The remaining reasons why queries returned zero results are as follows (with counts in parentheses): need spell-checker (16), need spell-checker for proper names (12), and the query assumes the corpus runs to the present time (e.g., “Hurricane Katrina” and “Obama election”, 7).

REFERENCES

- [1] Christel, M. G. and Frisch, M. H. 2008. Evaluating the contributions of video representation for a life oral history collection. In *Proc. ACM/IEEE-CS JCDL (Pittsburgh PA)*. JCDL '08. ACM, New York, NY, 241-250.
- [2] Shatford, S. 1986. Analyzing the subject of a picture: a theoretical approach. *Cataloguing Classif. Q.* 6, 3, 39-62.