

**Making History Useful: The Long Tail, the Search Economy and the
*U.S. Latino & Latina World War II Oral History Project Web Site***

BY

Dr. J. Richard Stevens
Assistant Professor
Division of Journalism
Southern Methodist University
P.O. Box 750113
Dallas, TX 75275
214-768-1915
stevensr@smu.edu

Dr. Maggie Rivas-Rodriguez
Associate Professor
School of Journalism
The University of Texas at Austin
1 University Station A1000
Austin, Texas, 78712
512-471-0405
mrivas@mail.utexas.edu

Abstract

This article presents two cases involving personally relevant searches that led users to the U.S. Latino & Latina World War II Oral History Project Web site. Using the long tail and the search economy paradigms (core components of current Web 2.0 development), the authors argue that the relevant searches were only possible because of the open archives of the site, and that the practice of denying open and free access to content by online news media affects their relevance to Web users.

Key Words: Web 2.0, long tail, search economy, archives

Introduction

Though online journalism remains in its infancy (Foust, 2005), development of its forms and structure are beginning to emerge. All media forms must adapt to their environment or die (Fidler, 1995), and how online journalism adapts to the ways in which Web users create relevancy is a critical issue facing news organizations.

Aizu (1997, p. 473) argued that the Internet provides a global arena for “minor culture” that “the mass media and or the mass economy cannot pay attention to.” These observations are consistent with the niche element of the long tail paradigm. The democratization of production tools and increasingly universalized access has allowed a greater variety and volume of content to exist on the Web than in other media. But within this arena of an ever-increasing amount of information, how are readers driven by particular interests able to find content that’s relevant? This question speaks to the heart of the supposed “Web 2.0” phenomenon: relevancy may be the key to the future development of the Web.

This paper looks at two cases involving personally relevant searches that led users to the *U.S. Latino & Latina World War II Oral History Project* Web site, which has displayed stories about its interview subjects since 1999. These relevant searches were possible only because of the open archive presented by the site, and demonstrate the principles of the long tail and search culture paradigms.

The Long Tail, Search Culture and Personal Relevance

The long tail is a popular colloquial term used to describe the economic shift in our culture away from mass culture to the adaptive personalization (“niche culture”) of goods and services. The paradigm is based upon a traditional feature of statistical distributions in which the vast majority of events or results in the distribution occur after the initial concentration of events. In other words, though the largest grouping of events occur near the beginning of the distribution, the remaining sparsely distributed events will number greater when aggregated than the largest clusters of events.

Although the term “long tail” is common in statistical texts, the popular use was first used by *Wired* editor Chris Anderson (2004) in an article describing the success of new media firms like Amazon.com and Netflix. In this article, Anderson described mass culture as a function of shelf space economics: the cost of stocking inventory in physical space drives brick and mortar firms to focus their production and sales on high volume sales items, keeping more eclectic media titles from being cost effective in the retail environment.

By contrast, online firms that can centralize their distribution centers are able to aggregate broadly dispersed demand into viable returns, and Anderson repeatedly demonstrates that the sum of these smaller niche demands over time exceed the total sales of the “hit” titles, which are prone to generate primarily short-term profit. Saul Hansell of the *New York Times* illustrated this principle when considering the life of a non-bestselling book:

The average book may sit on a shelf for six months or a year before it is bought. The cost of this inventory in a chain of hundreds or thousands of stores is huge. Amazon can keep just one or two copies in its warehouse – and still make the copy available to the whole country – and restock as quickly as customers buy books (Hansell, 2002).

When storage and distribution costs are low, more variety can be made available because of the decreased risks of overhead costs. Even the largest brick-and-mortar stores like Wal-Mart offer up to six times the number of available products on their Web sites (Thomas, 2002).

The second facet of this emerging economic paradigm is the rise of the “search culture” or “search economy.” The ability to easily navigate through hierarchies of choices has allowed for the commercial exploitation of the long tail. Battelle (2005, p. 155) illustrates this with the model of over-sized shoe vendors. In a particular market, the demand for over-sized shoes (much less for a particular size) will be small. But the “geography-busting” characteristics of the Internet make it profitable to sell several thousand shoes a month to the portions of markets world-wide that need oversized shoes.

Brynjolfsson, Hu and Smith (2003) presented a framework that sought to measure the effect of electronic inventory interfaces on consumer purchasing. They found that limited shelf space constrained the number and variety of products available for consumer choice when compared to the same vendors’ Web sites.

Together, the long tail and search economy constructs have allowed both researchers and professional writers to describe how companies like Netflix and Amazon are achieving more growth by offering a greater variety of products than traditional retailers. Netflix is estimated to offer 18 times the movie titles as Blockbuster and Amazon is estimated to offer almost 40 times as many books as a Borders Superstore (Anderson, 2006, p. 169).

Since the appearance of Anderson’s article and the numerous posts on his “The Long Tail” Web log, dozens of other applications of this theoretical concept have been

presented in popular media to describe related behavior. Many of these applications (such as online dating sites, social networking and consumer content sites) have been described as examples of “Web 2.0” activity. The Web 2.0 phrase was reported coined at a conference brainstorming session between O'Reilly Media and MediaLive International (O'Reilly, 2005). Though there has been some debate about precisely what Web 2.0 encompasses, most discussants agree that the features common to Google, Amazon, Myspace, Facebook and online dating sites: ways of building relevancy in order to increase access to materials with a greater amount of significance to the user.

Relevance has been an area of interdisciplinary study for years. In one of the earliest definitions of relevance, Rees (1966) defined it as “the criterion used to quantify the phenomenon involved when individuals judge the relationship, utility, importance, degree of match, fit proximity, appropriateness, closeness, pertinence, value or bearing of documents or document representations to an information requirement, need, question, statement, description of research, treatment, etc.”

As American society moves further away from mass culture and becomes increasingly fragmented and individualized, understanding personal relevance will become an increasingly important topic for researchers to pursue if they are to understand how users select and use online material. Thus, the distribution and access models embedded within Web 2.0 (and its underlying frameworks of the long tail and search economy) will have implications for online journalism.

Media and the Long Tail

Other content industries have used distribution and access models to explain the changing practices of their fields. Epstein (2005) discusses how technological innovation, in the form of DVD distribution, has changed the metrics that motion picture studios use to determine success and failure on the screen. The portability and easy access of the DVD medium has created a secondary use for motion picture content and the studios have taken notice and changed their production and tracking practices as a result.

In the business world, the key factors that determine whether a sales distribution can have a long tail are the costs of distribution and inventory storage. When distribution and inventory costs are high, popularity of the offerings is critical, for there will come a threshold beyond which the inventory costs exceed the return on sales. When distribution and inventory costs are low, selling less popular goods over longer periods of time becomes economically viable.

For online journalism outlets, the costs of storage and distribution are also extremely low when compared to traditional media. As media outlets store their older content in archives, these archives evolve into a virtual treasure chest of information and reference materials.

Using the Internet as a distribution method has allowed other types of niche media formats, such as podcasting, to develop. Podcasting is the automated download of digital audio content from one or more subscribed channels for future use. Because the audience for these subscribed content files may be spread across a larger geographic region than traditional radio distribution will allow, the aggregate audience for a program or segment need not represent a significant portion of a particular market.

When content is distributed through podcasting and other Web 2.0 delivery systems, the removal of geographic limitations in the distribution channel allow for a greater variety in available programming, as producers need not cater to the specific demographics of a particular market in their geographic region. This opportunity for niche programming reflects the same aggregation of demand outside of the traditional concentration of demand distribution as the long tail phenomenon

Recently, Good (2005) illustrated an unlikely example of this phenomenon. After languishing in obscurity for more than 200 years, William Playfair's *Commercial and*

Political Atlas and Statistical Breviary was recently brought back into circulation by information design guru Edward Tufte. Tufte, who frequently lectures about the history of information design has praised Playfair's book on several occasions, and because of this endorsement, the book has returned to print to capitalize on the demands of those interested in information design. Because of the low costs of distribution, storage and promotion, the book will turn a profit for firms like Amazon even if only a few thousand copies are ever sold, and even if no more than a single copy is ever sold in a given locale.

Though in less dramatic fashion than Playfair's book sales, this study demonstrates that allowing easy access to archival content is an important part of increasing the relevancy of journalistic content. In this manner, the long tail and the search economy can help online journalism outlets make content more relevant and thus drive more traffic to their sites.

U.S. Latino & Latina World War II Oral History Project

The *U.S. Latinos and Latinas & World War II* Web site grew out of the U.S. Latinos and Latinas World War II Oral History Project at the University of Texas at Austin School of Journalism. This project, begun in the spring of 1999, collects videotaped oral narratives of Latinos of all ethnicities of the WWII generation. The project interviews men and women in the military, as well as those in civilian life. In the fall of 1999, the project began publication of *Narratives*, a newspaper dedicated to the interviews. Interviews conducted by students enrolled in a class dedicated to the project and the occasional paid interviewer were used as the basis for feature stories. Stories from that first issue were posted on the project Web site shortly after the end of the

semester. Subsequent issues were published each semester in hard copy as well as posted to the project Web site.¹

Until 2003, the newspaper was mailed to approximately 3,500 individuals, institutions and organizations. The recipients were selected because they were involved in oral history, Latino studies or community affairs, journalism, or journalism education. approximately 1,000 more were distributed at conventions of oral historians, journalism educators and Hispanic journalists.

The mailings were curtailed in the fall of 2003 when the project could no longer support mailing costs. The newspaper continued to be mailed to approximately 300 recipients. The paper also was distributed free of charge at various conventions of journalists, oral historians, journalism educators and Latino organizations.

The newspaper's reach, limited to people, groups and organizations within the U.S., never surpassed 5,000 per issue. However, the stories from every issue were posted to the project's Web site, and were linked to and discussed online with increasing frequency.

The project eventually began to receive communication indicating that more and more interested parties were locating and using the *Narratives* content for a variety of alternative purposes (i.e., to raise cultural awareness, as classroom content, to enhance their veteran archives). Some made rather personal uses of the content, uses that would have been impossible had the stories remained exclusively in print.

¹ Both of the authors were involved in this project. (Full disclosure in printed version).

Connecting Families Across the WWW

Among the many different uses of media content occurred cases that exemplified personal relevance searches. Two respondents reported using the *Narratives* content to locate and reunite with family members. They attributed their ability to make that connection to the project site. In both cases, the offspring of WWII veterans were able to connect with their biological fathers, or their father's families, for the first time. In one case, the offspring was a woman in Belgium, in the other, it was a man in San Francisco.

Before stumbling upon the stories about the two veterans, neither party was aware of either the newspaper *Narratives* or of the oral history project that hosted the site. Both were simply conducting Web searches, looking for the specific names of the WWII veteran in question.²

Luis Leyva's Legacy

In the spring 2003 edition of *Narratives*, the project produced a story about Luis Leyva, a Mexican citizen who grew up in Laredo, Texas. Leyva had volunteered to serve in the U.S. Army during World War II (Flores, 2003). The article was posted to the project's Web site along with the remaining stories.

On April 29, 2004, the project director received an email from David Chaussée, a Belgian national, who requested contact information for Leyva or his children in order to pass along photographs of the veteran to the family. Chaussée, the president of a World

² Interview subjects signed a permission form permitting their information to be published in a variety of ways, as well as archived at a campus library eventually. Interview subjects were and are asked to approve the stories before publication. However, when requested, the information was not used either in the newspaper or on the Web.

War II history association, had discovered the Leyva story through a search engine and used the contact links on the site's contact page.

The project director forwarded the email to Leyva's daughter, Mary Alice Carnes, of Austin, Texas, who had interviewed her father for the project. Leyva had passed away in 2002. Carnes responded to Chaussée's request for contact the following day.

Chaussée explained that Leyva had been sheltered by a Belgian family during the war, before he and his unit were called into action. Chaussée had been commissioned to use his expertise to help a woman track down Leyva, and had found the *Narratives* archive online, which contained Leyva's story.

Chaussée proceeded to disclose to Carnes that her father had maintained a brief romance with a Belgium nurse and that the relationship had produced a child after his unit departed. Learning that she had a half-sister, Ms. Carnes corresponded with her new-found family and in May 2006 traveled to Belgium to reunite with a previously unknown branch of her family tree.

Although some of her immediate family members were cautious about the new additions to the family, Ms. Carnes was happy: "I've already accepted them in my heart as family and they've accepted all of us, now we just need to get to know one another (Carnes, 2006)."

Joe Longoria Meets His Son

Another story featured in the spring 2003 edition of *Narratives* produced similar results. The online story about Joe Longoria, a Mexican-American Army cook who was

captured after the Battle of the Bulge and held as a prisoner-of-war, caught the attention of John Garcia.³ Garcia emailed the project manager of the Web site and as a result of another forwarded email, met his biological father for the first time in January of 2006.

Garcia said later that he had not believed his father was alive:

When I was a young boy, my mother told me that my father had perished in WWII, since she had never heard from him after he shipped out. They were not married when I was born. My mother told me that my father's name was Joe Longoria, that he and his bother Daniel Longoria were both stationed in Ft. Bliss, Texas, during war. (Garcia, 2006b)

After his mother's death, Garcia began to search through the Army's casualty records, looking for information about his father. When he found the records of Daniel Longoria, but not Joe, he decided to broaden his search.

Googling his father's name, Garcia stumbled upon the *Narratives* article. Using this information, he consulted *Ancestry.com* and began to piece together his family's story. Soon he had enough to embark on a fulfilling journey. Afterwards, he sent the project an email:

All my life (60 plus years) I believed my father had perished in WWII, based on what my mother had told me. When I read the article in the oral history project, it confirmed that he was in fact alive. Without your help it would have been more difficult for me to locate my father. I met him last weekend. It was a wonderful and moving experience." (Garcia, 2006a)

Garcia also noted that his father was suffering from dementia and that he did not believe Longoria understood their relationship. He was introduced to the elderly man as a "friend of the family." (Garcia, 2006b)⁴

³ "John Garcia" is a pseudonym used at the request of the respondent. The name of his father here is also a pseudonym, as it appears he is unable to give informed consent to use his real name at this time of this writing.

⁴ The Longoria family also asked that the story be pulled from the Web after Mr. Garcia came forward. They said that it was not completely understood that the reach of the Web was not entirely understood previously. The project complied.

Discussion

In his book, Anderson (2006, p. 54-56) proposed three “ascendant forces” that allow firms and individuals to make use of long tails:

- 1) democratizing tools of production,
- 2) lowering the transaction costs of consumption, and
- 3) connecting consumers to drive demand to niches.

All of these factors contributed to the connections made between the individuals mentioned in this article and the Web site. Because the technology used in the *Narratives* publication and presentation, many individuals were able to contribute their time and energy to the project without having to be in physical proximity with the building in which it was produced. Volunteers and family members were able to contribute background information, photographs and leads for other stories via email, a necessity for a project covering so many different communities across the nation. The democratization of tools allows a greater degree of access for those who wish to lend their effort.

The cost of mailing the publication swiftly became one of the larger barriers to raising awareness about the project. Posting the material allowed for the content to be available to more people in more areas of the world than the print publication had an opportunity to reach.

The third force, Anderson’s version of the search economy, is perhaps the most powerful and the most interesting aspect of the events under review.

The Long Tail of Narratives

At the height of its production, a single issue of *Narratives* reached less than 5,000 individuals. No data exists for the number of people who have read *Narratives* online, but

the varying uses and broader reach demonstrated that many more people have read the content on the Web than in newsprint.

Two of Anderson's major principles of the long tail are the recognition that more aggregate consumption will occur over time than in the initial "buzz" period (in this case, publication of a new edition) and that this recognition justifies the production and marketing of niche products.

Had the project stopped at the print publication of the oral history stories, many readers would never have encountered the stories. And as the long tail continues to progress, it is quite likely that more readers of a given *Narratives* story will have consumed the content in online form, and long after the printed edition it originally appeared in has disappeared from circulation.

The other principle Anderson cites is the niche concern. When the cost of production is high, producers of content are encouraged to produce output that will have mass appeal, thus justifying the investment needed for the production. But when the cost of production is low, producers of content can focus their energies on "minor culture" projects with more focused appeal. The cost of producing and storing the *Narratives* content on the Web is negligible, allowing project to continue producing these stories with the knowledge that the content will be used in time whether the printed publication has a broad reach or not. This also allows the project to add content to the archives, making the *Narrative* stories available for years to come.

Once again, the force that enables this tail is the search economy.

Finding Narratives

The search economy's value to the two cases reviewed is immeasurable. Simply put, without the ability to easily search and find content on the Web, neither the Belgian history association president nor the out-of-state son could have ever encountered the *Narratives* content they needed to find the individuals they sought.

To date, the *U.S. Latino & Latina World War II Oral History Project* Web site has not used any search metatags, nor has it been actively registered with any search database. The attention and readership it draws has been primarily the result of passive indexing methods and social referrals.⁵

Posting the *Narratives* content on the WWW allows those who seek content about Latinos or war veterans to find places where those two topics intersect. As the requests for information, contributions and messages referring to additional uses continue to arrive, the *Narratives* content is proving most popular and more useful in its electronic form.

Neither of the unusual stories cited in this study would have been possible without the search economy and the long tail that it utilizes. In each case, the searcher did not come across the content until over a year after the print version of the publication had been published. Nor was either searcher aware of the *Narratives* publication prior to their search.

⁵ As of the fall of 2006, the site is undergoing a wholesale redesign by *Utopia*, UT's digital libraries site, which will include metatags.

Accessible Archives: The Key to Online Journalism?

Recently, Tremayne (2004) found that the number of links per journalism story and the increased variance in story topic has increased over the past five years. These findings do suggest the growing “web of context” that he (and others) cite.

Clearly, the low cost of production and the search economy suggest that more of journalistic content will be consumed after the “breaking news” stage of delivery through Web 2.0 use (i.e., driven by personal relevance) than during the “breaking news” stage. As online journalism outlets strive to become relevant to their readership, context and the ability to allow easy navigation of relevant content would seem to be an important part of utilizing the long tail.

However, many journalism sites have begun to close off their archives to the public, allowing a 14-day window for free access to content and requiring a subscription or per-use fee for access to older material.

This practice would seem to run counter to the principles of the long tail and the search economy. As users increasingly search the Web for content that is relevant to their lives, paid archives (or a lack of archiving itself) can create barriers to online readership trends.

Battle (2005) illustrated this problem when he described why the *Wall Street Journal* and the *Economist* were beginning to fall off the reading lists of online users.

Because they are fearful of losing revenue as a result of search, both require paid subscriptions, and therefore, neither supports the kind of deep linking that drives news stories to the top of search results at Google and its brethren. In other words, both are very difficult to find if you get your daily dose of news, analysis, and opinion from the Internet. And as we all know, folks who read their news on newsprint ain't getting any younger. (Battelle, 2005, p. 173).

Tremayne's findings concerning the increased number of links in online journalism stories are encouraging. However, if the majority of these links are locked behind

archives (or soon to be locked off), it does not seem that these links will encourage the amount of traffic needed to improve readership and search engine rankings.

Online journalism has encountered a new content sharing paradigm. Althaus and Tewksbury (2002) found that importance and prominence of stories was obscured in the online edition of the *New York Times* when compared with the print edition. This implies that users are more likely to be driven by personal relevance when selecting news content from the online edition.

In order to drive the amounts of links and traffic needed to support online advertising models, more working links need to be available for users to find the relevance necessary to connect. Bloggers and other sites will not generally provide links that don't work. "In today's ecosystem of news, the greatest sin is to cut oneself off from the conversation. Both *The Economist* and the *Wall Street Journal* have done just that." (Battle, 2005, p. 174).

The *U.S. Latino & Latina World War II Oral History Project* Web site does not initiate any Web 2.0 techniques to promote or improve its online popularity. And yet two people searching for specific information were able to find their way to the online publication without prior knowledge of its existence. This was only possible because the archives are open to the public and permanently hosted to a linkable address.

As Scheufele (2001) observed, the relationship between mere media consumption and mobilization relies on interpersonal discussion. In this case, being discussed and linked to by other discussants is the only reason why the *Narratives* content appears so prominently in search engine results.

Granted, the content of the site and the eventual use of the content was extremely niche, but that niche made it extremely relevant and popular to those who were interested. The print industry practices about providing access to earlier content are based on the assumptions of high costs of storage and distribution. Newsprint is heavy and bulky to store and index.

By not adopting the print industry mentality of making older content more difficult to find, the *Narratives* content is assured a much greater audience in the long tail than it ever would have had in the “short tail” of print publication.

Sources

- Aizu, Izumi. (1997). In *The Harvard Conference on The Internet And Society*. O'Reilly and Associates.
- Althaus, Scott L. and David Tewksbury. (2002). Agenda setting and the "new" news: Patterns of issue importance among readers of the paper and online versions of the New York Times. *Communication Research*, 29, 180-207.
- Anderson, Chris. (2004). The long tail. *Wired Magazine*. (October): 170-177.
- Anderson, Chris. (2005). "A 4th long tail Business Category: Tools," *The long tail: A Public Diary on the Way to a Book*, (July 28), available online at http://www.thelongtail.com/the_long_tail/2005/07/a_4th_long_tail.html.
- Anderson, Chris. (2006). *The long tail: Why the Future of Business is Selling Less of More*. New York: Hyperion.
- Battelle, J. (2005). *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York: Portfolio.
- Brynjolfsson, E., Hu, Y., and Smith, M.D. (2003). Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Book Sellers. *Management Science*, 49 (11), 1580-1596.
- Carnes, Mary Alice. (2006). Email conversation with the author. (March 28).
- Epstein, E. J. (2005). *The Big Picture: The New Logic of Money and Power in Hollywood*. New York: Random House.
- Fidler, Roger. (1995). *Mediamorphosis: Understanding New Media*. Thousand Oaks, CA: Pine Forge Press.
- Flores, Monica. (2003). "I Know No Other Country." *U.S. Latinos and Latinas & World War II*, 4 (1), available online at <http://utopia.utexas.edu/explore/latino/narratives/v4n1index.html>
- Foust, James C. *Online Journalism: Principles and Practices of News for the Web*. Scottsdale, AZ: Holcomb Hathaway.
- Garcia, John. (2006a). Email conversation with the author (January 28).
- Garcia, John. (2006b). Email conversation with the author. (March 30).
- Good, R. (2005). Information Design Classic Makes The Shelves Again After Two Hundred Years. *MasterNewMedia* (September 27), available online

at

http://www.masternewmedia.org/news/2005/09/27/information_design_classic_makes_the.htm

Hansell, Saul. (2002). News Analysis: Amazon looks for an encore. *New York Times*. (January 26): C2.

Hiam, Shelley. (2003). Army Cook Survived Months as Prisoner of War in Germany. *U.S. Latinos and Latinas & World War II*, 4 (1), available online at http://utopia.utexas.edu/explore/latino/narratives/07Dimas_Pete.html

O'Reilly, Tim. (2005). "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software." *O'Reilly*, available online at: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

Rees, A. M. (1966). The relevance of relevance to the testing and evaluation of document retrieval systems. *Aslib Proceedings*, 18(11), 316-324.

Scheufele, Dietram A. (2001). "Democracy For Some? How Political Talk Both informs and Polarizes the Electorate." *Communication and U.S. Elections: New Agendas*. Lanham, MD: Rowman and Littlefield: 19-32.

Thomas, Owen. (2002). Lord of the Things. *Business 2.0* (March): 32.

Tremayne, Mark. (2004). "The Web of Context: Applying Network Theory to the Use of Hyperlinks in Journalism on the Web," *Journalism and Mass Communication Quarterly* 81:2 (Summer): 237-253.