Large U.S. Newspapers Update Their Websites More Often

by Jin Xu

An analysis of the news websites of 59 large, medium and small U.S. newspapers revealed that almost 40 percent of all stories were updated with large newspapers setting the pace for the most immediacy in their lead stories.

American newspapers are shifting away from a static product on paper. By delivering a dynamic flux of continuous information on their websites, they have been evolving into an open, dynamic service that breaks the time paradigm of the 24-hour news cycle.1 The change has redefined the practice of newspapers in reporting lead news stories as well as keeping them updated. Emphasis is placed on immediacy, one of the most fundamental changes that online capability brings to the news business.2

Tremayne, Weiss and Alves studied 24 American online newspapers and found a steady increase in dynamic journalism between 2004 and 2006.3 The present study builds on that. Its objective is to examine the life and updates in lead news stories. It is distinguished from the study by Tremayne and colleagues’ in several ways.

First, this study examined dynamic changes to individual stories. It focused on the most dynamic content of the site, the lead news story, to find out how frequently newspaper sites reported news and kept it updated as well as to determine how the updates were distributed.

In addition, it included the category of dailies with smaller print circulation so that newspapers of all sizes were represented.

Third, the study continuously monitored the lead news story on the websites of 59 newspapers and collected real-time dynamic content changes at 10-minute intervals.

Xu is an associate professor in the Department of Communication Studies at Winona State University.
intervals, 24 hours a day, seven days a week. Dynamic content is defined as story content that changes during its life as the lead news story.

Results obtained may advance the theoretical understanding of how the Web has transformed news production. Specifically, they shed light on how the lead news story responds to today’s twenty-four-hour news environment. Thus, the study contributes to mass communication theory-building in the Internet age.

Literature Review

Contemplating the effect of new media on news, Pavlik suggested that as a real-time medium, the Web would alter journalism practices. The speed and computer format of real-time news delivery would have an effect on news writing and news reporting practices. Stories would be shorter and would build in several takes. They would be disseminated immediately. An example is reporting real-time financial news. This is in sharp contrast to the traditional practice of integrating information into a single story.

New media have made news content ubiquitous. Audiences have access to more news and information as news sources are increasingly viewing themselves as content providers. In addition, many online news ventures original to the Internet have been able to take advantage of the capabilities offered by new media. News aggregators collect and organize news stories from a variety of sources and update them constantly. All of these developments have put pressure on news organizations to keep their news updated in a timely fashion.

On March 17, 2009, the Seattle Post-Intelligencer terminated its print publication. While print newspaper readership is declining, online readership is on the rise, as online editions extend beyond the reach of print newspapers. According to The Project for Excellence in Journalism, online readership was up substantially, pointing to a clear direction of news consumption. News consumers are changing their expectations and behaviors. Seeking out information on demand, they turn to outlets that can tell them what they want to know when they want to know it.

News organizations have been adapting themselves to these changed expectations and behaviors. News is shifting from being a product to becoming a service. According to The Project for Excellence in Journalism, as news consumption becomes continual, more new effort is put into producing incremental updates. Newspapers have improved their websites with a 24/7 diet of breaking news. Because immediacy is among the most important driving forces for the site’s popularity, the improvements build visitor traffic and induce at least some readers to spend more time on the site. As a result, some newspapers have become major players in traffic. For example, The New York Times has been consistently ranked among top current events websites at Nielsen/NetRatings.

Research Questions

The lead news story is the most dynamic content on newspaper sites, which post the newest lead story as soon as possible and add expanded material where and when possible. These practices suggest that three dimensions of the lead story may be examined as indicators of dynamic content, story life, update and growth. When news breaks, a fresh lead story is published. The publication pushes the old story out. Story life measures how long a story stays as the lead news story.

RQ1: What is lead news story life in American online newspapers?

Another indicator is update. To examine its frequency and distribution patterns, it is further differentiated into four sub-indicators: update count, interval, immediacy and clustering. While update count measures how many times a story is updated, update interval measures the average time between updates.

RQ2: What is lead story update count in American online newspapers?

RQ3: What is lead story update interval in American online newspapers?

No matter how frequently the story is updated, updates are not made randomly. There must be a distribution pattern. Discovering it adds to the understanding of online news consumption as well as production. For example, it helps explain the findings that media dependency increases during breaking news events. It also helps predict what this dependency means to news consumers.

Update immediacy and update clustering were used to examine the distribution pattern. Update immediacy measures how soon updates are likely to be made after the publication of the original story. To measure update immediacy, story life was divided into four stages of equal length—stages 1, 2, 3 and 4—each representing a quarter of the story life. The earlier the stage, the younger the story life. Update immediacy reveals whether updates are more likely in younger or older aged stories. In comparison, update clustering indicates how soon another update is likely to be made after the current update is posted. The more updates that are made sooner, the more clustered the updates are. Update clustering is distinguished from update immediacy in that the former is a measurement of time lapse between two adjacent updates while the latter is a measurement of time lapse between the story’s original publication and the current update.
RQ4: What is lead story update immediacy in American online newspapers?

RQ5: What is lead story update clustering in American online newspapers?

Existing research indicates that live reporting affects story length in television news. This may also be the case with news on the Web.

Online breaking stories are very iterative—you print what you know when you know it and then you add to it.

The additions may include latest developments and background information to help the user make sense of what has happened. In both cases, they contribute to story growth.

RQ6: What is lead story growth in American online newspapers?

Tremayne and colleagues studied how dynamic journalism rose on the Web sites between 2004 and 2006. They found that small papers were catching up, showing a significant rise in dynamic content. Research question seven looks into a similar comparison on the story level.

RQ7: How is newspaper size related to lead story life, update and growth?

By using content changes in certain types of news as indicators, Tremayne and colleagues found a significant rise in dynamic content from 2004 to 2006. Where was the trend headed five years later? Research question nine answers the question.

RQ8: How are categories of proximity related to lead story life, update and growth?

Based on the characteristics revealed in this study, it is concluded that immediacy is a noticeable hallmark of the lead news stories on American newspaper sites.

Method

Unit of Analysis

The unit of analysis is the lead news story, defined as the story visually established by a single large photograph and the largest headline on the newspaper’s homepage. An individual story is identified by its URL address instead of the title because modifications to the title may be considered an update. For example, the story “Obama Signs Healthcare Fixe’s Bill” was treated as an update to the story “Obama Set to Sign Healthcare Fixes’ Bill” if they shared the same URL address.

Sampling and Samples

Newspaper size was determined by the circulation of the daily newspaper. Altogether, 60 newspapers were originally selected from the Editor & Publisher International Yearbook 2007. The large newspaper group was made up of the top twenty newspapers on the book’s list of the Top 100 Daily Newspapers by Circulation. Twenty mid-size newspapers were randomly selected from the remaining eighty newspapers on the list. Twenty small newspapers were randomly selected from the yearbook. To examine the dynamic content of frequently updated stories, it is crucial to collect the stories continuously at regular intervals for a reasonable period.
of time. In this research, the data collection was conducted by a networked computer once every 10 minutes, 24/7 for three randomly selected weeks in three years, June 14 through 20 in 2009, Aug. 1 through 7 in 2010 and March 20 through 26 in 2011.

**Data Processing, Coding and Analysis**

The sampled story was cleaned to obtain the pure news text, which includes the title, byline, story body and copyright lines. The cleaned texts were grouped by their URL addresses, producing 4,029 unique stories. As each story was iteratively accessed every ten minutes, there were repetitive copies, which were discarded to obtain all its updated copies. A copy was deemed repetitive if no modification was detected when compared with the copy immediately preceding it. Modifications include addition or deletion of number(s), word(s), sentence(s) or paragraph(s), as well as changes of order in them.

Story life, update count, update interval and story growth were calculated. All of the stories were coded on categories of proximity. If the URL address provided proximity information, coding was done as the URL indicated. For example, if a character string like /local/ or /world/ was found, the story was coded as local/regional or national/international accordingly. If there was no such indication, story content was examined to determine the geographic focus. The authorship of the story was also used to assist the determination. The data were entered into SPSS, where independent sample t-test and one-way ANOVA were performed.

**Findings**

The results show that 38.4 percent of the stories were updated. The average story life was 347.8 minutes, 388.1 minutes for updated stories and 324.2 for un-updated stories (t(4027)=4.39, p<.001). Update rates for large, mid-size and small papers were 52, 33.9 and 22.3 percent respectively. Updated stories averaged 662.1 words per story compared to un-updated stories’ 585.7 words (t(4027)=4.75, p<.001). The following results are on the updated stories. The story life, update count, update interval and story growth are summarized in Table 1, answering research questions 1, 2, 3 and 6.

**RQ4 asks about the distribution pattern of update immediacy.**

The results show that the younger a lead story was, the more likely it was to be updated. As it aged, its chances of getting updated dropped gradually, but at the end, there was a surge in updates. As is shown in Figure 1, the slope of the curve decreased for the most part and then increased. Specifically, the odds of a story getting updated in stage 1 were 33 percent. The probability dropped to 25.9 in stage 2, 22 in stage 3 and 19 in stage 4.

**RQ5 asks about the distribution pattern of update clustering.**
Because of the 10-minute interval in data collection, update clustering was examined as the percent of updates made within the intervals of 10, 20 and 30 minutes. The results show that updates were likely to appear in clusters. The tendency suggests that the posting of an update made it more likely for another update to appear sooner rather than later. Figure 2 shows that 42.6 percent of the updates were made within 10-minute intervals. The rate dropped sharply to 16.7 percent within 20-minute intervals and to 9.6 within 30-minute intervals. Together, 69 percent of the updates were made one after another, no longer than 30 minutes apart.

RQ7 asks how paper size is associated with lead story life, update and growth.

ANOVA tests show differences between large, mid-size and small papers in story life (F(2,1546)=18.5, p<.001), update count (F(2,1546)=6.9, p<.005) and update interval (F(2,1546)=17.3, p<.001). These differences came from comparisons between large papers and mid-size or small papers. Large papers had a shorter story life and were updated at shorter intervals than mid-size or small papers. In update count, there was a difference between large and mid-size papers but not between large and small papers. The tests show no differences between mid-size and small papers. The differences in story growth were not statistically significant (F(2,1546)=1.17, p=.31) among large, mid-size and small papers. Differences were observed in update immediacy. More updates were likely in stage 1 for large papers than mid-size papers and for mid-size papers than small papers. [See Table 2] For large and mid-size papers, update rates dropped in stages 2 and 3 but reversed the downward trend in stage 4. For small papers, the update rate was highest in stage 1 and stage 4. Updates clustered more in large papers than mid-size papers and in mid-size papers than small papers. Table 3 shows that 46, 38 and 36 percent of the updates were made to large, mid-size and small papers, respectively, in 10-minute intervals. The rate dropped sharply in 20-minute intervals and continued the steep decline in 30-minute intervals. Altogether, updates made one after another no longer than 30 minutes apart added up to 71, 64 and 70 percent for large, mid-size and small papers.

RQ8 asks how lead story life, update and growth are associated with proximity.

T-tests show statistically significant differences in story life, t(1511)=2.29, p<.05 and update interval, t(1511)=2.68, p<.05. There was no difference in update count, t(1511)=23, p=.82 and story growth, t(1511)=1.48, p=.14. There was a small difference in update immediacy. National or international news was more likely to be updated in its earlier stages than were local or regional news. Table 2 shows that 33 percent of the updates were made to national or international news compared to 30 for local and regional news in stage 1. Again, update rates dropped in stages 2 and 3, but rose in stage 4. There was a bigger difference in update clustering. Table 3 shows that with 45, 17 and 10 percent in 10, 20 and 30 minute intervals, updates to local and regional news clustered more than those to national and international news, whose corresponding figures were 38, 16 and 9 percent. Altogether, updates made one after another no longer than 30 minutes apart added up to 72 percent in local and regional news and 64 percent in national and international news.

RQ9 asks whether lead story life, update and growth changed from 2009 to 2011.

The answers are positive. As shown in Table 1, the trend was toward a shorter

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Update Immediacy and Paper Size, Geographic Focus and Year</strong></td>
</tr>
<tr>
<td><strong>Paper Size</strong></td>
</tr>
<tr>
<td>Large paper</td>
</tr>
<tr>
<td>Mid-size paper</td>
</tr>
<tr>
<td>Small paper</td>
</tr>
<tr>
<td><strong>Geographic Focus</strong></td>
</tr>
<tr>
<td>National or intern*tal news</td>
</tr>
<tr>
<td>Local or regional news</td>
</tr>
<tr>
<td><strong>Year</strong></td>
</tr>
<tr>
<td>2009</td>
</tr>
<tr>
<td>2010</td>
</tr>
<tr>
<td>2011</td>
</tr>
</tbody>
</table>

*Indicating fraction of updates made in the stage.
Table 3
Update Clustering and Paper Size, Geographic Focus, and Year

<table>
<thead>
<tr>
<th>Paper Size</th>
<th>Interval between Updates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10 minute</td>
</tr>
<tr>
<td>Large paper</td>
<td>.46</td>
</tr>
<tr>
<td>Mid-size paper</td>
<td>.38</td>
</tr>
<tr>
<td>Small paper</td>
<td>.36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geographic Focus</th>
<th>Interval between Updates</th>
</tr>
</thead>
<tbody>
<tr>
<td>National or intern’tal news</td>
<td>.38</td>
</tr>
<tr>
<td>Local or regional news</td>
<td>.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Interval between Updates</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>.30</td>
</tr>
<tr>
<td>2010</td>
<td>.45</td>
</tr>
<tr>
<td>2011</td>
<td>.46</td>
</tr>
</tbody>
</table>

*Indicating fraction of updates made in the interval.

Discussion

The online world is a highly competitive market environment, which brings out more dynamic content. Based on the characteristics revealed in this study, it is concluded that immediacy is a noticeable hallmark of the lead news stories on American newspaper sites. A shortened news cycle, update immediacy and update clustering contribute to the immediacy.

First, the news cycle was shortened as indicated by story life, update count and update interval. Take large newspapers, for example. With an average story life of 277.2 minutes, they put out 5.19 lead stories a day. Fifty-two percent of the stories were updated, averaging 4.9 updates per story. The updates grew the story by 25.6 percent. This is a significant improvement on the traditional 24-hour print news cycle.

Second, immediacy resulted from the manner updates were made. More updates tended to come out at a story’s early life. The most recent stories are always the most frequently updated, which feeds news consumers with timely information. Consequently, for those who are keen on breaking news and eager to learn the latest developments, it is meaningful to go back frequently to the site. This helps explain why media dependency increases during breaking news events.25
Finally, updates tended to group in clusters. Clustering significantly shortens the intervals between adjacent updates in a cluster, providing a reason for news consumers to check back frequently for timely information when they see a recently posted update. The findings on update clustering raise a question about what message an update ticker, such as “Updated 5 minutes ago” gets across. There is no doubt that it suggests current information. However, according to update clustering, it has two layers of meanings. In addition to assurance of fresh information, it should also be an indication that another update is coming soon or an invitation to check back soon. Do news consumers take it as such? More study is needed.

The surge in updates toward the end of a story life seems contradictory to what update immediacy suggests. It is also counter-intuitive that more frequent updates are made to an aged story before it is replaced. If stories are seen as discrete units independent of each other, the surge does not make much sense. However, many stories are inter-connected, covering continual developments of a news event. In this context, update clustering and update immediacy shed light on the surge. Suppose that instead of a series of inter-connected stories, there is only one single overarching story, spanning a few days. According to update clustering, the story is interspersed by periodic clusters of higher-frequency updates. In a certain cluster, an update may be significant enough to transform the story, giving birth to a new story. Because the birth occurs in a cluster, more frequent updates are found both before and after it. Updates before the birth show up as the surge, whereas updates after the birth are consistent with what update immediacy predicts. The birth of a new story is analogous to the breakout of an earthquake. A surge in seismic activities (small shocks) leads to a major earthquake, followed by frequent smaller aftershocks, which taper off until the next surge.

The unique contribution of this study is to investigate dynamic journalism on the individual story level. It has made a couple of other findings. First, it shows that local and regional news have caught up with national and international news in dynamic content. While differences in story life and update interval exist, there is little difference in update count and story growth. Second, with its more recent evidence from story level comparison, instead of page level, the study partially supports Tremayne’s and colleagues’ claim that smaller papers were catching up with large and mid-size papers in dynamic content. Table 1 shows that as far as updated stories are concerned, small papers have caught up with mid-size papers, putting out as many lead news stories and updating them just as frequently.

However, the results show that both small and mid-size papers were still behind large papers in dynamic content. Three possible explanations are offered below. First, it might be attributed to the fact that small and mid-size papers are likely to focus more on local or regional affairs, which are less fluid than national or international events. As Table 1 shows, national or international news stories have a shorter story life and are updated at shorter intervals.

Second, it might be attributed to the use of wire service feeds, characteristic of immediacy. As large newspapers report more national or international stories, they are more likely to use news feeds from wire services, resulting in more dynamic content. This possibility was investigated by examining the use of wire service feeds. The finding is inconclusive. The results show that large papers used more wire service feeds than did mid-size papers, 17.6 percent against 10.1 percent. This seems to strengthen the argument that the use of wire service feeds adds to dynamic news. However, compared with small papers, which used 27.7 percent of wire service feeds, large papers used fewer feeds but still lead in dynamic content production. It seems that the use of wire service feeds did not necessarily increase dynamic content. The third explanation is that small and mid-size papers have fewer resources than do large papers, which limits their ability in dynamic news production. The findings show that large papers updated a larger portion of lead news stories than did mid-size and small papers. It appears that resources are associated with the number of updates. More studies in this area also are needed.

Limitations and Future Research

A limitation of the study is that the time dimension of the life and update results are close approximation to actual numbers. This is because of the tradeoff 10-minute data collection interval. Reporting/updating breaking news is a rolling process. It can occur at any moment. No matter how short the sample collection interval is, there is an error of measurement. Sampling the lead news story at a ten-minute interval may not capture every single update, nor does it capture the story right at the moment it is posted. Setting shorter intervals, such as one minute in length, may make the approximation closer to the actual number, but it is still an approximation. The actual number is hard to obtain. Moreover, shorter intervals may complicate data collection and lower data accuracy because of network outage, heavy traffic to the sites, etc.

This study identified key indicators of immediacy and showed that differences exist in the lead news stories when newspaper sizes, geographic focuses and publication years were compared. It used large samples to examine these differences at required significance levels. Large samples, however, might compromise the predictive power of statistical models. Future study topics could include the examination of other key explanatory factors that account for the differences, such as topics and the disruptiveness of the news, and the development of predictive models for the dynamics of lead news stories using such explanatory factors.

Researchers may also look into the individual updates and categorize them in terms of content updates and mechanical changes, such as correcting typographical errors. Considering that a large number of updates were made within short time periods, it would be interesting to find out what caused these modifications. It is worthwhile to examine the number and types of error cor-
rections as well as content updates in terms of background information and depth of coverage. Studies in this area shed light on how the intense pressure to publish quickly affects the long-standing importance of traditional journalistic values, such as accuracy, clarity, good grammar and spelling.  

Conclusion

Throughout newspaper history, timeliness has been defined in the context of creating a daily product. But the Web has changed this. By monitoring the dynamic content changes in the lead news stories of 59 American newspaper sites 24/7, this research has produced empirical findings about their online news cycle concerning paper size, a story’s geographic focus and change over time. In addition to story life, update count and update interval, it has identified update immediacy and update clustering as indicators of immediacy. The knowledge of online news contributes to the development of mass communication theories in the Internet age, and this research opens a new area for more research into the news on the web.

Notes

21. Tremayne et al., “From Product to Service.”
22. Ibid.
23. Ibid.
25. Wu and Bechtel, “Web Site Use.”
28. Wu and Bechtel, “Web Site Use.”