

Improving the Online News Experience

V.G.Vinod Vydiswaran
University of Illinois
Urbana, IL
vgvinodv@illinois.edu

Raman Chandrasekar
Microsoft Research
Redmond, WA
ramanc@microsoft.com

ABSTRACT

News consumption patterns are changing, but the tools to view news are dominated by portal and search approaches. We suggest using a mix of search, visualization, natural language processing and machine learning to provide a more captivating, sticky news consumption experience. In this position paper, we suggest a design for one specific use-case where a user needs to catch up on news from a particular time period. The results need to cover key events that happened during the time period, but the stories should be prioritized based on the user's interests. Further, users should be able to interact and explore stories of interest. We present the limitations of existing online news sites for such a task and present some ways to address these issues.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *information filtering, selection process.*

General Terms

Algorithms, Human Factors.

Keywords

News summarization, news interfaces, news personalization

1. INTRODUCTION

The news landscape has undergone major changes with the advent of online media. While the readership of traditional newspapers has declined over the past few years, the consumption of news over the Internet has increased significantly. As with other kinds of online information, the dominant mode of assessing news online is through search. According to the Pew research conducted over Apr–Jun 2008 [16], 83% of those going online for news use search engines to find stories of interest. So, even though most search engines (and other sources) have dedicated portals for news, consumption of news is triggered primarily through queries. Search engines today address this user behavior by integrating relevant news results with Web search results for news-related queries.

However, this mode of presentation of news is not optimal. Even as news is shifting online, the presentation of news is still driven by the print media. There is limited real estate on the search result pages to display news, and many news articles do not get surfaced on the site. Finding relevant news is more than just retrieving news results or restricting the search based on keyword queries over the news domain. Presentation of news needs to cater to specific user needs. We propose a use-case driven approach to selecting relevant news stories and presenting these appropriately to the user. Further, users should be able to explore the news landscape – getting to other related news articles, visualizing the connections between stories, getting background information on relevant people and concepts, commenting on and annotating stories, and sharing interesting items with friends.

In this position paper, we survey the existing online news sites in Section 2 and outline their limitations. In Section 3, we present a specific use-case for news consumption, and suggest a design that meets the requirements of this use-case. We conclude with related work and potential directions.

2. SURVEY OF ONLINE NEWS SITES

In this section, we briefly survey different online news sites, and present a summary of some of their features.

2.1 News Aggregators

Traditional search companies form the largest and most frequently visited online news portals. According to the most recent Pew Research biennial news consumption survey [16] conducted in 2008, Yahoo! (28%) and MSN (19%) are the most frequented websites among Web news users, ahead of traditional media outlets, such as CNN (17%). Google was polled at 11%. Other news ranking sites, such as *Digg* [5] and *Reddit* [17], have grown in popularity, but only 5% of Internet news consumers use these to find news stories.

Traditional search engines aggregate news from many sources and categorize them by news categories, such as World, Business, Sports, etc. Aggregating news from multiple sources helps them present a multimodal view of a news story that includes videos, photos, and live feeds. Sites such as *Yahoo! News* [26] prioritize news sources and differentiate news coming from different sources. Users can select the source per category, and the news stories from that source are displayed. Other sites, such as *Bing News* [1] and *Google News* [8], present news stories as clusters across multiple sources. This allows them to also incorporate stories from other sources, such as blogs, *Twitter* [23], and *Wikipedia* [24].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HCIR'10, August 22, 2010, New Brunswick, NJ, USA.

Copyright 2010 ACM 1-58113-000-0/00/0010...\$10.00.

2.2 Traditional News Media Online

While readership and viewership of traditional newspapers and television news diminished over the past decade, their online versions have seen an increase. The average monthly reach of web newspapers among Internet households has increased from 27.4% in 2004 to 40.9% in 2008 [15]. Many television channels now make news clips available online either on their sites or on other video-sharing sites like *YouTube* [27]. Newspapers have augmented their online content with videos and photos to visually appeal to younger readers on the Web (for example, see Figure 1). With the news consumption moving away from print media, some newspapers, such as the *Seattle Post-Intelligencer* [18], have gone web-only, while other news outlets, such as *The Huffington Post* [21], *Newser* [12], and *Seven-Sided Cube* [19], present editorial and blog content as independent news online. In addition, a large section of news is community-generated. Sites such as *Newsvine* [14] and *GlobalReporter* [7] allow users to post (and rate) news local to their community. Other news media outlets, such as *CNN iReport* [3], have also accepted this notion of grass-root journalism and allow users to post news videos.



Figure 1. Print (left) and online (right) versions of *The Washington Post* [22] on June 10th, 2010. Note the integration of (a) video, (b) images, (c) live market feeds, and (d) social networking sites in the online version.

2.3 News On-the-fly

News consumption patterns have also changed over the past decade. Rather than setting apart time to access news, consumption is spread throughout the day. In a recent survey, 73% of all online users say they come across news online when they have been on the Web for another purpose. This is especially true of younger news consumers, who typically follow links to news stories, rather than go to news sources themselves [16].

Social networking sites have played an important role in pushing news content to the web users. Sites such as *StumbleUpon* [20], *Digg* [5], *Reddit* [17], *Yahoo! Buzz* [25], *Delicious* [4], *Facebook* [6], and *Twitter* [23] allow users to tag news stories and recommend or share them with friends on their social network. Such news recommendation services not only rate mainstream news, but also help users stumble upon unique news stories that they may not otherwise have a chance to see. Many traditional news sites have interfaces (plug-ins) to these bookmarking sites to enable readers to share news freely. The bookmarking also helps present “popular” or “upcoming” news, as shown by news ranking sites like *NewsPulse* [13], *BuzzFeed* [2], *Digg*, and *Reddit*.

2.4 Analysis of Online News

As illustrated above, many online news sites augment news with videos, images, and user comments to enrich the news consumption experience. Search-based news aggregator sites cluster and categorize news stories and allow users to customize and personalize what they want to read. Services such as email/mobile news alerts and RSS feeds, and customized web pages, such as *My MSN* [10], *My Yahoo!* [11], and *iGoogle* [9] allows users to get news on demand. Readers are encouraged to share news with others, either in their social network or the online community at large. News ranking sites use these to assess popularity of sites and surface articles that are generating a lot of interest (“buzz”).

The algorithmic aggregation of news across sources seems to treat all news sources equally, especially when selecting which news item to show. In addition, recent updates often supersede earlier reports, even if the earlier reports were from reputed sources. However, users may prefer local news sources for local news; or in general, specific news sources for different genres of news. Local news or news from a particular region is often under-represented, because there are fewer sources reporting on regional news. Some sites have space set apart for local news, but this tends to be limited.

The “one size fits all” approach of using keyword query based retrieval is not optimal for news. Query-based triggering is often imperfect, and searching for news using just keyword queries often limits expressivity. News demands a different ranking than web search. News dissemination is more than just selecting a list of news articles about popular events from well-known news sources. Online news must instead cater to specific use-cases and should ideally be personalized to the user.

3. News Sync: CATCHING UP ON NEWS

In this section, we present a specific scenario for a user-driven news digest to illustrate our ideas. We propose techniques to select and present news according to user needs and preferences. While the techniques used may not be new, we suggest that the integration we propose will lead to a better news experience.

3.1 Motivation

Consider the following scenario: Katie is an avid news reader who tracks news on a daily basis, often following up on specific news events several times a day. Sometimes Katie may be cut off from news sources, for example, when she goes on a long vacation. When she is back online, she may want to know what happened while she was away. She may want to skim through the major news stories that took place, including updates on the news she was following regularly before going on vacation.

This caters to a common, specific need of a news consumer wanting to catch up on news. The scenario becomes more compelling if Katie migrates to another city or country and loses touch with traditional sources of news. Katie continues to be interested in local news from her place of origin. However, unless she is willing to visit every site regularly, she may not get the news of interest.

We propose a system we call *News Sync*, which allows Katie and similar news consumers to get adaptive, personalized news digests covering a period of time, a region, a topic or a combination of these.

3.2 Requirements for *News Sync*

We list the following requirements for *News Sync*:

1. **Control over news categories, topics, and sources:** The user should be able to specify the time period of interest. In addition, the user may specify if she is interested in news from particular sources, specific news categories, locations/regions, and/or specific topics.
2. **Personalized news feed:** The system should figure out which stories are currently the most relevant to the user, based on past user behavior and user preferences, similar in spirit to work by Billsus and Pazzani [29].
3. **Variety in news content:** The system should show a variety of content across diverse categories, instead of, say, returning a list of ten “most popular” news links which may be restricted to one or two topics. Users can thus get an overall picture of key events first, before they delve into specific stories.
4. **Adaptive and integrated news presentation:** The news interface needs to be adaptive to the category of news and presence of multiple modes of news content. For example, news about Harry Potter over summer 2007 should include, among other stories, the trailers from the movie “Harry Potter and the Order of Phoenix” (video), book reviews of “Harry Potter and the Deathly Hallows” (text, blogs) – which were both released in July 2007 – along with pictures and news about the Harry Potter theme park announced in May 2007 (images).
5. **Interactive and exploratory user interface:** The user should be able to interactively and directly modify time, location, and other news parameters and have the system respond immediately with updated views of relevant news. Further, the system should support browsing related news articles.
6. **Parameterized interface design:** Users should be able to set system parameters to get results at different specificities. For example, a user might ask for news about a specific topic (e.g. “Top Kill” in context of efforts to reduce the oil spill in Gulf of Mexico), or news that is of interest to the user based on search history (e.g. other efforts by BP to reduce the oil spill). The user may also request other related news based on her profile (e.g. news about impact of the oil spill on the marine environment, on the livelihood of people along the Gulf coast, on tourism, or on BP’s share price).
7. **Support source-tracing and finding related news:** The system should allow users to go from any news summaries to the original news articles. Further, the system should suggest other related news articles based on the news items viewed.
8. **Ability to share news:** Users should be able to comment on and share interesting news articles over their social network.
9. **Support news analyses by sentiment and points of view:** Users should be able to view stories pivoted/summarized on sentiment or different points of view.

3.3 Key steps in building *News Sync*

We are building a prototype system based on the requirements listed in Section 3.2. To achieve these goals, we propose the following steps:

1. **Collecting a news corpus:** Our first step is to get indexed access to the news articles for the time period of interest to the user population, along with source, location, and date information. In addition, articles are processed with a named entity recognizer, to identify key concepts.
2. **Selecting key news stories:** News stories are then selected based on a number of features, including the content, topic trends, the number of sources covering the news, the number of articles on the news story, the volume of news content, and various aspects of the user model (user profile, explicit user preferences, and implicit interest tracking). The selection criteria also include the time-spread of the news to identify key events that the user should know about.
3. **Selecting relevant threads:** Once the key stories are identified, the relevant individual articles need to be retrieved. This involves filtering based on time and location, and incorporates some of the model preferences described above. The system may choose to ignore or prefer news sources based on the user preference model and granularity of news. Local news sources may be preferred for local news stories, while national mainstream news sources may be preferred for broader scenarios.
4. **Summarizing news stories:** News needs to be presented in a manner that is easy to consume. This involves selecting what content to present and how best to present it. This may involve adaptive summarization across documents based on the user model, and presenting diverse points of view.
5. **Presenting and visualizing news:** News stories may be visualized on a timeline, a map, etc.

Once such a system is developed, it can be evaluated using implicit/explicit feedback and through a survey to understand the usage patterns and the features that are popular.

Such a system would help frequent travellers, business customers who need to know the impact of ongoing news on their business, and avid news followers who spend the most time with the news.

We have developed a prototype of the *News Sync* system as part of this year’s HCIR Challenge.

4. RELATED WORK

Past literature has looked into generating a personalized webpage of news relevant to the user based on the topics of interest. Kamba, et al. [31] did one of the early studies on presenting an interactive newspaper on the Web. They propose a system that builds web pages dynamically as the user browses the newspaper. Anderson and Horvitz [28] developed a personalized web page as a montage of links of frequently viewed pages that changes dynamically with the time at which the page is viewed. The system learns which pages are viewed regularly at certain time periods and presents content based on the user interest and browsing pattern. For example, a user might be shown weather forecasts and key news in the morning; the stock price ticker and work-related resources during the day; and traffic pattern and TV listings in the evening. In our system, we propose to follow the

users' interests to prioritize the news presented, and also allow users to specify topics or other inputs such as time and location of interest.

There has also been work in providing personalized newsfeeds. Gabrilovitch, et al. [30] analyze inter- and intra-document differences and similarities to recognize novel content in articles and how the information has evolved over time. This helps them develop measures to rank news by novelty, and pick the best (most novel) update to send to the user as a newsfeed. Other researchers, such as Tintarev and Masthoff [34] have studied other measures of similarity of news headlines to improve news recommendation.

There is a lot of relevant work in the realm of interface design. For example, Shneiderman [32] suggests use of dynamic queries to update the search results as users adjust sliders and other UI elements. Teitler, et al. [33] suggest NewsStand, which proposes using geographic information in news articles to overlay news on a map. This presents users with a geographic perspective of where the news comes from and helps them cluster and explore news based on location.

Some news ranking sites are able to show "popular" news for particular days or months, based on how many users clicked on or shared a news article. Figure 2 shows an example that displays the top 10 *digg* articles for the month of April, 2010. Displays such as these allow for simple presentation of high level summaries.

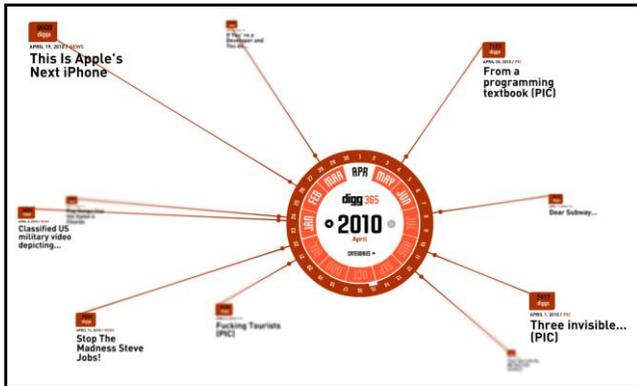


Figure 2. Top 10 *digg* articles in April 2010. The larger the font size, the more *diggs* the article received.

5. CONCLUSION

In this paper, we propose an approach to providing a captivating, sticky news consumption experience, using techniques from search, language processing, visualization and learning. We listed requirements for a specific use-case of catching up on news. We propose a design for a system to address this use-case. We are working on such a prototype and hope to have the system ready to demo at the workshop.

6. REFERENCES

[1] Bing News. <http://www.bing.com/news>
 [2] BuzzFeed. <http://www.buzzfeed.com/>
 [3] CNN iReport. <http://www.ireport.com/>
 [4] Delicious. <http://delicious.com/>
 [5] Digg. <http://digg.com/>

[6] Facebook. <http://www.facebook.com/>
 [7] Global Reporter. <http://globalreporter.com/>
 [8] Google News. <http://news.google.com/>
 [9] iGoogle. <http://www.google.com/ig>
 [10] My MSN. <http://my.msn.com/>
 [11] My Yahoo!. <http://my.yahoo.com/>
 [12] Newser. <http://www.newser.com/>
 [13] NewsPulse. <http://newspulse.cnn.com/>
 [14] NewsVine. <http://www.newsvine.com/>
 [15] Nielsen/Net Ratings, MegaPanel Data. <http://www.naa.org/TrendsandNumbers/Newspaper-Websites.aspx>
 [16] Pew Research Center for the People and the Press, 2008. Key News Audiences Now Blend Online and Traditional Sources – Audience Segments in a Changing News Environment. *Pew Research Center Biennial News Consumption Survey*. <http://people-press.org/reports/444.pdf>
 [17] Reddit. <http://www.reddit.com/>
 [18] Seattle Post-Intelligencer. <http://www.seattlepi.com/>
 [19] Seven-Sided Cube. <http://www.sevensidedcube.net/>
 [20] StumbleUpon. <http://www.stumbleupon.com/>
 [21] The Huffington Post. <http://www.huffingtonpost.com/>
 [22] The Washington Post. <http://www.washingtonpost.com/>
 [23] Twitter. <http://twitter.com/>
 [24] Wikipedia. http://en.wikipedia.org/wiki/Main_Page
 [25] Yahoo! Buzz. <http://buzz.yahoo.com/>
 [26] Yahoo! News. <http://news.yahoo.com/>
 [27] YouTube. <http://www.youtube.com/news>
 [28] Corin R. Anderson and Eric Horvitz. 2002. Web Montage: A Dynamic Personalized Start Page. In *WWW*.
 [29] Daniel Billsus and Michael J. Pazzani. 2000. User modeling for adaptive news access. In *User Modeling and User-Adapted Interaction*, 10(2) pp. 147-180.
 [30] Evgeniy Gabrilovich, Susan Dumais, and Eric Horvitz. 2004. NewsJunkie: Providing Personalized Newsfeed via Analysis of Information Novelty. In *WWW*.
 [31] Tomonari Kamba, Krishna Bharat, and Michael C. Albers. 1993. The Krakatoa Chronicle - An Interactive, Personalized, Newspaper on the Web. In *WWW*.
 [32] Ben Shneiderman. 1994. Dynamic Queries for Visual Information Seeking. In *IEEE Software*, 11(6), 70-77.
 [33] Benjamin E. Teitler, Michael D. Lieberman, Daniele Panozzo, Jagan Sankaranarayanan, Hanan Samet, and Jon Sperlberg. 2008. NewsStand: A New View on News. In *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS)*.
 [34] Nava Tintarev and Judith Masthoff. 2006. Similarity of news recommender systems. In *Workshop on Recommender Systems and Intelligent User Interfaces*, G. Uchiyigit (Ed).