NARC: The News Article Revision Comparator

Aaron St. Clair, Matthew Fong, Nicholas Diakopoulos, Irfan Essa
School of Interactive Computing | Georgia Institute of Technology
85 5th St. NW, Atlanta, GA 30332-0760, USA
astclair@gatech.edu, fong@gatech.edu, nad@cc.gatech.edu, irfan@cc.gatech.edu

ABSTRACT

Currency of information in news consumption is an important facet of information quality which involves both the journalist providing updated information and the consumer being *aware* of updates and changes to the news stream. We are addressing information quality and currency in online news articles from the viewpoint of news consumption with the intent of reducing the consumption effort involved in getting the most up-to-date information on a breaking news story. The goal of this research is thus to develop a web-based user interface which (1) allows users to easily and quickly see updates to news articles online and (2) blends into existing consumption patterns by integrating into news websites. We have built NARC to address these issues by providing an integrated interface which allows users to quickly perceive changes to news articles using an inline text visualization.

Keywords: news, information quality, visualization

INTRODUCTION

Anyone who has watched CNN on a busy news day knows that breaking news items are constantly being updated with the latest information that comes in from reporters, interviews, video cameras, and increasingly from ordinary citizens snapping and sending in photos. The problem is that even though new information is constantly being added and updated, we still have to sit through and consume the information that we are already familiar with. This observation has inspired us to build a system that can visually signal where news information has been updated or renewed so that a news consumer can choose to focus on what is truly current and new in a news story.

Currency is a component of *information quality*, which itself is studied in information science as a composition of the accuracy, comprehensiveness, validity, reliability, and currency or recency of information [4]. Much of the news industry is predicated on ensuring the timely delivery of *quality* information, embodied in the journalistic principles of accuracy, fairness, thoroughness, and trustworthiness [3]. Journalists have the difficult task of balancing timely delivery of current information with the other components of quality such as accuracy and reliability which may require time to guarantee.

From the point of view of the news consumer, having the most recent or current information is desirable since revisions to the information stream by the journalist may serve to enhance the quality of a story by either (1) adding new

information, (2) updating information which has changed, or (3) removing information that may have been shown to be unreliable or inaccurate after initial publication. We maintain that the nature and sequence of changes to a news article also contains information which may be valuable to a news consumer.

We have built The News Article Revision Comparator (NARC) system to address the information quality component of currency in news consumption. By highlighting the most recent updates to an article we hope to reduce the user burden of focusing on news updates and to facilitate better information quality by providing transparency to the process of news article updates by journalists. NARC is both a server which tracks changes to online news articles on BBC News and a lightweight browser-based client interface and visualization which allows the user to see news article changes from the last version of the story they visited. The interface allows for the fluid exploration of the sequence of changes made to the news article. Our visualization provides an improvement over existing textual comparison visualizations by showing insertions and deletions to the current article inline rather than by using a side-by-side layout and is integrated into the BBC News site interface to provide a seamless user experience.

SYSTEM DESCRIPTION

Architecture and Implementation

The NARC system makes use of the Revisionista [1] site which tracks changes to articles on the BBC News site. These revisions are supplied as an RSS feed which the NARC server parses every 10 minutes. Each revision is then fed into a server-side database which stores the article text, timestamp, and version information.

On the client side, NARC is implemented as a Greasemonkey plugin. Greasemonkey is a scripting plugin for the Firefox browser which allows 3rd parties to install scripts (similar to javascript) which run on the client when particular websites are visited. Once a user has installed the

Figure 1. The NARC script injects interface elements into the webpage of a BBC article.

Figure 2. Visualization showing both insertions (green highlight) and deletions (light gray strike-through) from a previous version of the article.
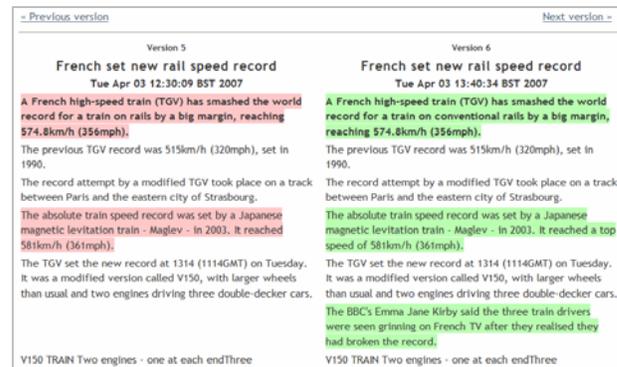

Figure 3. A text diff visualization from Revisionista showing versions side by side. Recognizing changes requires reading back and forth between versions.

NARC greasemonkey script, it is executed every time the user visits a BBC News article webpage.

The script inserts some additional interface elements into the page (see Figure 1 at top) and executes a text diff algorithm which recognizes insertions and deletions between the article that the user is currently viewing and any of the revisions that are stored in the database for that article. The insertions and deletions are then rendered on the page using a combination of javascript and CSS properties.

Interface and Tasks

We envision a scenario of *breaking news* consumption where a user may visit a news site several times in the course of a day in order to get the latest information and updates about a rapidly evolving story. The most important things for the user to see are thus the insertions (new information) and deletions (old information potentially being replaced) without having to read through material that they may already be familiar with.

The NARC interface is designed to be lightweight and available if the user is interested in exploring (Figure 1). If a user is uninterested in insertions or deletions they can be toggled using checkboxes. The slider bar corresponds to a timeline with each tick representing a version of the article. We chose a slider because of the natural timeline metaphor this provides in representing a time series of articles. The user can compare the current version of the article to any other historical version of the article by simply dragging the handle of the slider. The slider defaults to compare against the most recent prior version of the article. Textual feedback is also given to indicate the version and time-stamp of the article being compared against.

Visualization

We strove for a parsimonious visualization of insertions and deletions that would quickly convey the changes between two versions of an article (see Figure 2). The visualization provided by Revisionista places two article versions side by side and highlights changes in red and green (Figure 3). We wanted to get away from the concept of the eye having to multiplex between two display spaces in order to integrate what the effective changes to the text were. This led us to integrate the insertion and deletion visuals into *one combined view* of the current article compared to a revision. While there are many different rendering styles that could be adopted, we believe the current visuals quickly highlight newly inserted information (using light green) and provide access to older information (using light gray strike-through) without the older information being too

distracting. These renderings will be tested for readability when we evaluate the system.

CONCLUSIONS AND FUTURE WORK

NARC is a system which addresses information quality and currency in online news articles from the viewpoint of news consumption with the goal of reducing the consumption effort involved in getting the most up-to-date information on a breaking news story. Our interface makes it easy for a news reader to see the latest updates to a news article directly in their browser and integrated into their existing news consumption routine. The visuals we have created simplify the perception of changes to articles by reducing the effort needed to spatially multiplex the eye between different versions.

There are some interface refinements and extensions that can be incorporated into NARC including facilitating other types of tasks beyond consumption. For example a media analyst might be interested in analyzing a news story over a large period of time or in comparing any of the interstitial revisions of an article. We would also like to explore the use of animation in the visualization such as in [2] to be able to cue people to changes in the text of the article using kinetic typography or to be able to track changes using an animated transition between versions. We intend to do a user study to explore the usability of the interface and the readability of the text renderings. We are also interested in extending the system to incorporate news sources other than the BBC.

REFERENCES
1. Revisionista. http://www.newssniffer.co.uk/articles/list_by_revision (Accessed June 18, 2007)
2. Baudisch, P., Tan, D., Collomb, M., Robbins, D., Hinckley, K., Agrawala, M., Zhao, S. and Ramos, G., Phosphor: explaining transitions in the user interface using afterglow effects. in *Proceedings of UIST* (2006), ACM, 169-178.
3. Kovach, B. and Rosenstiel, T. *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect*. Three Rivers Press, 2001.
4. Taylor, R. *Value-Added Processes in Information Systems*. Ablex, 1986.