News Rover: Exploring Topical Structures and Serendipity in Heterogeneous Multimedia News

Hongzhi Li*, Brendan Jou*, Joseph G. Ellis*, Dan Morozoff* and Shih-Fu Chang
Digital Video & Multimedia Lab, Columbia University

ABSTRACT

News stories are rarely understood in isolation. Every story is driven by key entities that give the story its context. Persons, places, times, and several surrounding topics can often succinctly represent a news event, but are only useful if they can be both identified and linked together. We introduce a novel architecture called News Rover for re-bundling broadcast video news, online articles, and Twitter content. The system utilizes these many multimodal sources to link and organize content by topics, events, persons and time. We present two intuitive interfaces for navigating content by topics and their related news events as well as serendipitously learning about a news topic. These two interfaces trade-off between user-controlled and serendipitous exploration of news while retaining the story context. The novelty of our work includes the linking of multi-source, multimodal news content to extracted entities and topical structures for contextual understanding, and visualized in intuitive active and passive interfaces.

1. INTRODUCTION

The rise of digital media and the Internet has had unprecedented impact in journalism, and on both news curators and news consumers alike. With regard to consumption, there has been steady decline of viewership in broadcast television (TV) news as consumers migrate to Web-based platforms to find articles, shorter news clips, and blog posts. The on-demand nature, flexibility for mobile consumption, and closer-to-realtime updates is believed by many consumers to triumph waiting for a newspaper print or scheduled, lengthy video broadcast. As this slow transition widens the gap between broadcast and online content, consumers are disengaging with the rich video media available through broadcast in favor of less-informative, but bite-sized news, and news curators are faced with challenge of giving stories the necessary context with short-form media, e.g. a 140-character blog post.

Despite the declining popularity of the traditional 30-minute or 60-minute news program, the throughput of news content both broadcast and online has only exponentially grown. An ongoing challenge in the research community to address the data explosion has been to develop high-throughput, content-based indexing and annotation solutions. However, these media are often studied in a single-source domain setting. Only recently are some works investigating heterogeneous indexing from multiple sources.

Our system, called News Rover, integrates multiple news sources and the multi-view nature of news, and offers an immersive user experience in a semi- and fully-automatic setting. The key novel contribution of our system architecture is its ability to link and index content from heterogeneous news sources, including broadcast TV news, online articles, and social media feeds (like Twitter), organizing them into a topical structure for greater story context. Our web interface can be visited at `http://ptvn.demo.dvmm.org`.

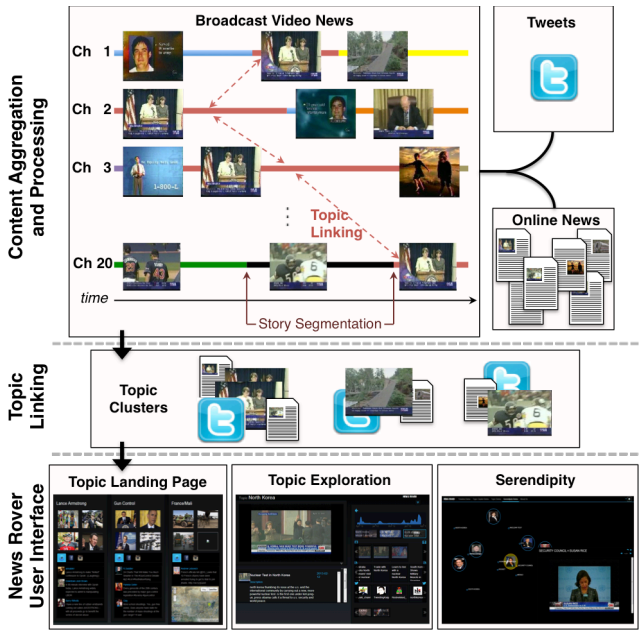

Figure 1: System Architecture Overview.

2. CONTENT AGGREGATION

Recording and Crawling. Our recording architecture consists of 12 cable TV tuners equipped with on-chip encoding. The system records continuously from a mix of analog and digital channels as transport stream media, generating about 700 hours per week and covering about 100 channels.

*Denotes equal contribution.

MM '13 Barcelona, Catalunya Spain

Table 1: Average Estimates of Data Per Week.

Programs recorded	700	Online articles	72,000
Hours of video	700	Google topics	4,000
Stories segmented	7,000	Twitter trends	3,500

We implemented a scheduling routine that queries an electronic program guide feed every hour for an up-to-date listing of news programs. The system schedules a recording job for the next available TV tuner for every newly discovered news program.

In addition, we developed a crawler that queries Google News every five minutes for new unseen topics and downloads all articles indexed under that topic, together with long-term tracking of 2,000 topics. A Twitter crawler was also developed to grab trending topics and all the tweets indexed under that trend. We also track and crawl about 2,000 long-term hot Twitter trends every day.

Story Segmentation. Recorded TV programs consist of a single contiguous video consisting of several stories. Our system automatically cuts each program into several story segments. In many U.S. programs, closed captions (CC) contain >>> characters to demarcate the beginning of a new topic. To overcome the time lag of CC, we apply a dynamic time warping algorithm to align the CC text to automatically recognized speech. To further refine the time precision, we performed shot detection [3] and chose shot boundaries to be likely story boundaries based on proximity with >>> symbols and shot lengths. Segmenting a program into stories is the most costly process in our pipeline, taking ∼20 minutes for a 60-minute program. In the case when CC story markers are unavailable, we will apply our prior results on multimodal story segmentation [2], which demonstrated satisfactory performance with F1 score up to 0.76.

Topic Linking. Given that the data crawled from Google News and Twitter are organized into topics, we required a cross-source listing of trending topics to resolve differences in semantic tags. To do this, we linked related Google News topics and Twitter topics using both visual and textual content cues. Visual matching was accomplished by finding near-duplicate image pairs in images from online news articles and tweets. The number of near-duplicate image pairs was applied as the visual similarity metric. For text matching, we used a TF-IDF model to calculate the similarity between two topics. A combination of visual and textual cues is used to get the final similarity topic-pair score. A similar method was used to also link segmented video news stories to trending topics.

Entity Extraction. The CC text in TV news is a rich source of annotation on the video, containing explicit mentions of persons, locations and organizations associated with a story. Since CC is caseless, we then performed named entity recognition using a model trained on sentence examples where every letter was capitalized [1]. Since we had aligned the CC to the speech, time stamps were also assigned to detected named entities, specifying exactly when they were mentioned in the video. Since named entities extracted from CC are not always accurate, we used DBpedia and Wikipedia to correct and normalize them. This resolves typographic errors and aliases and also allows for profile pictures to be extracted. On average, we extract about five names per story segment.

In addition to named entities, we also applied a keyword extraction algorithm to extract the important concepts in each topic. Given the time occurrence of named entities, we find points in the video where a key concept and person is mentioned in CC within a 15-second time window of each other. This co-occurence mining of relations between entities-to-entities and entities-to-key-concepts allows us to explore how the "major players" shape the news. We are also currently developing a multimodal algorithm for extracting quotes associated with each of these named persons (i.e., extracting "who said what") to provide additional linking modes and measure an entity influence on a topic.

3. NOVEL VISUALIZATION INTERFACE

The complex space of multi-source, multimodal, multi-dimensional data is wrought with visualization challenges as humans simply cannot visualize their beyond certain dimensions. We designed a discovery visualization user interface (UI) allowing users to explore the space of linked, multi-source, multimodal news data. We accomplish this in two ways: a topic-organized shuffle exploration UI, called *semi-automatic*, and a physics-simulating bounce serendipity UI, called *fully-automatic operation*. We contrast this with a *manual* navigation of news which requires experts to pool sources and draw context themselves. Please see our supplementary material and online site for more detailed demonstrations.

Structured Exploration. To investigate the in-depth multimodal linking structure that would otherwise be unapparent, we focused our interface toward ease of use and concurrent discovery of these non-obvious links. Our current design consists of an interactive matrix display of multimodal data. Subliminal topic linking structure exists within the data, but the UI combines and reshuffles the representation on-the-fly to show related entities.

Serendipity. In many cases, news consumers are interested in a show-me-what-you-got paradigm that reveals surprising, compelling facets previously unknown. To address this, we exploit human intuitive understanding of the physical world for data discovery. Our current design utilizes non-deterministic modeling via gravitational simulation and elastic collisions to model the importance of players in shaping news events and the serendipitous interaction between players and concepts/events involved in a topic. This interface overcomes the linearity of reasoning and successfully adds new visualization paradigms that are particularly appealing to consumers preferring a passive lean-backward consumption approach.

4. REFERENCES

[1] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proc. Annual Meeting of ACL*, 2005.

[2] W. Hsu, L. Kennedy, C.-W. Huang, S.-F. Chang, C.-Y. Lin, and G. Iyengar. News video story segmentation using fusion of multi-level multi-modal features in TRECVID 2003. In *ICASSP*, 2004.

[3] J. Mathe. `http://johmathe.name/shotdetect.html`.