What Do They Think?
Aggregating Local Views about News Events and Topics

Jiahui Liu and Larry Birnbaum

Northwestern University

j-liu2@northwestern.edu, birnbaum@cs.northwestern.edu

ABSTRACT

The web has become an important medium for news delivery and consumption. Fresh content about a variety of topics and events is constantly being created and published on the web by many sources. As intuitively understood by readers, and studied in journalism, news articles produced by different social groups present different attitudes towards and interpretations of the same news issues. In this paper, we propose a new paradigm for aggregating news articles according to the news sources related to the stakeholders of the news issues. We implement this paradigm in a prototype system called LocalSavvy. The system provides users the capability to aggregate and browse various local views about the news issues in which they are interested.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information storage and retrieval-*search process*, H.4.m [**Information System Applications**]: *miscellaneous*

General Terms: Human Factors, Design, Experimentation

Keywords: News aggregation, local opinion

1. INTRODUCTION

The web has become an important medium for disseminating and consuming news. The wide variety of news content available on the web provides readers unparalleled opportunities for accessing news from around the world. On the other hand, the web presents a challenge in effectively navigating this massive amount of information. New techniques are required to find and aggregate relevant and interesting news content for users.

In this paper, we explore the meaningful differences in news sources to aggregate different points of view about news issues. While journalists often strive for objectivity, news production is nevertheless at least partly a subjective process affected by the cultural norms and values of the social groups that the authors represent [5]. For the same news issue, different social groups may have different points of view about that issue. Readers are implicitly aware of this, which is why they are often interested in knowing not only the facts (i.e. "X did Y"), but are also in the opinions of the actors (i.e. "What does X think about Y").

According to van Dijk [5], news articles from different social groups may be dissonant with the ideology held by readers, resulting in more interesting and memorable news reading. Finding the local points of view is also useful, for individuals, organizations and governments alike. For individuals, knowledge about other social groups can be illuminating and help them discover and overcome prejudices [4]. For policy makers of

organizations and governments, the opinions of their counterparts are helpful for making informed decisions.

In this paper, we propose a new paradigm for aggregating news articles according to the news sources associated with the *stakeholders*, i.e. person, organizations and locations involved in or potentially affected by the news issues. By aggregating news articles from various sources associated with the stakeholders of the news issue at hand, we can answer questions such as "Do the Iraqi people want U.S. troops in their soil?" or "How do the Chinese regard the human rights record of the United States?"

We implement the proposed paradigm in a prototype system called *LocalSavvy*, which analyzes the news article that the user is interested in and aggregates local opinions about the news issue. Our user study demonstrated that news articles aggregated by LocalSavvy presented relevant and distinct opinions of the stakeholders, which can be clearly perceived by the subjects.

2. LOCALSAVVY

LocalSavvy aggregates news articles according to the news sources associated with the stakeholders. The association between news sources and stakeholders is established through locations. Many of the entities in news articles are actually locations, such as cities and countries. The news articles published at the location provides the local views about the news issue. For other types of entities, such as people and organizations, we approximate their views with articles from the news sources that collocate with the entities. This approximation is based on the observation that local news sources usually speak in favor of the entities that are members of the local social group. Furthermore, we can distinguish an official news release from unofficial news sources. The distinction will provide insight into the differences between the governments' perspectives and those of the general public.

LocalSavvy is implemented as a server-based system. Users provide URLs of the news in which they are interested and LocalSavvy finds and aggregates news articles from the news sources associated with the stakeholders in the news event or topic. When LocalSavvy presents the news results to the readers, instead of showing a generic snippet of the news articles, the system extracts the local opinions from the news web pages and summarizes them in the results. Moreover, when a user clicks on one of the search results, the local opinions of the corresponding stakeholder are highlighted in the news web page.

3. IMPLEMENTATION

Figure 1 shows the architecture of LocalSavvy. There are mainly three steps: news story modeling, news aggregation, and opinion extraction. The three steps are supported by a commonsense knowledge base, which acquires its knowledge from the web.

The *News Story Modeler* retrieves the news web page that the user is interested in and creates a model about the news event or topic. To model the news story, the system creates two vector

presentations separately for named entities and non-named entity words. We adopted the modified TF-IDF term weighting presented in Liu et al. [3], which implements the insight that important information is presented near the top in news writing. The modified Term Frequency (TF) is calculated from both the frequency and position of term occurrence. The weight of each word is the sum of the scores for all of its occurrences, with word occurrences near the top of the article being assigned higher scores. The named entities are weighted with the modified TF. Other non-entity words are weighted with modified TF-IDF, with IDF computed from an archive of 343,187 news stories.

Based on the two vector representations built by the story modeler, the *News Aggregator* constructs queries by combining the top named entities and non-entity words from the two vectors. Besides query formulation, another key process of the aggregator is to identify the related locations of the news issue. The system collects the locations, people, and organizations as stakeholders. For people and organizations, LocalSavvy finds the related locations of the entities in its knowledge base. For instance, state officials are matched to the countries that they belong to. After location identification, the system sends the queries to Google News Search [2]. For official news releases, the search is restricted to the related government websites; for unofficial news articles, the search is restricted to news sources in the location.

After LocalSavvy finds the local news articles, the Opinion Extractor retrieves the web pages and identifies the opinions of the stakeholders. Firstly, opinion sentences are identified using subjective clues derived from FrameNet [1]. After opinion identification, the system selects the opinion holders from the persons and organizations within the opinion sentences with a set of extraction rules. Similar to news story modeling, the opinion holders are queried in the knowledge base to find the associated location. If the location matches the source location of the news article, the sentence is highlighted as a local opinion sentence.

LocalSavvy consists of a knowledge base to store the commonsense knowledge it needs for story modeling and searching. A distinctive feature of the knowledge base is that it autonomously acquires knowledge from the web. When an information request sent to the knowledge base is not satisfied, the system will automatically search the web to find the answer using a set of extraction rules and a voting mechanism similar to QA systems. The new knowledge is then cached for future use.

4. USER STUDY

We conducted a user study to evaluate the efficacy of the system in finding local opinions of news issues. To collect the test articles for the user study, we ran LocalSavvy on the "World News" RSS feed of Google News on October 15, 2007. Within the 20 news items in the RSS feed, for 9 items LocalSavvy was able to find local news articles from more than three stakeholders. The news items were presented to the subjects as *background story*. For the 9 news items, we gathered the top two search results for the top three stakeholders and used them as the *test articles*. Altogether we collected 50 test articles about 9 news issues.

22 students of our department participated in the study. The subjects chose two news issues in which they were interested and read the background stories. Then the subjects were asked to classify the test articles into one of the three stakeholders mentioned in the background story, according to the points of view expressed in the test articles. It should be mentioned that the
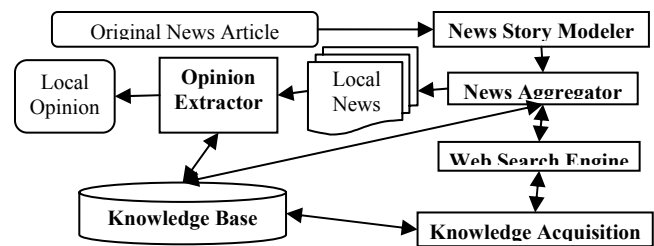

Figure 1. System architecture

test articles were presented to the subjects as plain text documents with only titles and content of the news article. The subjects had no knowledge about the sources of the test articles in the study.

If the test article represents the perspectives of the stakeholder strongly enough to be clearly perceived by the readers, the article should be consistently assigned to the correct stakeholder. Based on this idea, the system was evaluated by the classification accuracy of the subjects. In our study, the subjects generated ranked lists of classifications for the test articles, with highest rank of 2 and lowest rank of 0. For each test article, we computed its *degree of partiality* by averaging the ranks of the true classification (i.e. the real stakeholder) for that test article.
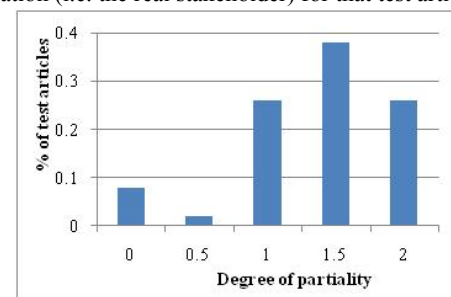

Figure 2. Distribution of degrees of partiality

Figure 2 illustrates the distribution of degrees of partiality for the 50 test articles. The mean degree of partiality for the 50 test articles is 1.51, with standard deviation of 0.52. As shown in figure 2, a large portion of the test articles express the opinions of the stakeholders clearly enough to be perceived by the readers.

5. CONCLUSION

In this paper, we propose a new paradigm for aggregating news articles according to the news sources associated with the stakeholders. We implement this paradigm in LocalSavvy, which helps users to aggregate and browse local opinions of news issues they are interested in. Our future work will focus on identifying and highlighting opinion differences expressed in various local news reports retrieved by LocalSavvy

6. REFERENCES

[1] Boas, H. C. "From Theory to Practice: Frame Semantics and the Design of FrameNet". In Semantisches Wissen im Lexikon. Tübingen: Narr. 2005.

[2] Google News. http://news.google.com/. 2007.

[3] Liu, J., Wagner, E., and Birnbaum L. "Compare&contrast: using the web to discover comparable cases for news stories". In Proc. of 16th WWW. 2007.

[4] Perry, D.K. "News reading, knowledge about, and attitudes toward foreign countries". Journalism Quarterly, 67(2). 1990

[5] van Dijk, T. A. "News as discourse". Hillsdale, NJ: Lawrence Erlbaum. 1988.