

---

# Automated News Generation for TV Program Ratings

**Soomin Kim**

Human-Computer Interaction &  
Design Lab.  
Seoul National University  
soominkim@snu.ac.kr

**JongHwan Oh**

Human-Computer Interaction &  
Design Lab.  
Seoul National University  
whee0501@snu.ac.kr

**Joonhwan Lee**

Human-Computer Interaction &  
Design Lab.  
Seoul National University  
joonhwan@snu.ac.kr

**Abstract**

Automated journalism, automatically generating stories based on algorithms, has received considerable critical attention in diverse fields. However, automated journalism has not addressed the TV industry in much detail. This research aims to create a system to automatically generate news about TV ratings. The framework will involve undergoing the processes of data gathering, identifying important events by predefined algorithms, generating a story in narrative format, and publishing the output. The algorithm that determines the structure of the stories is defined by analyzing existing news about TV ratings that reflects key variables. Although the output of the research is limited to one type of news template, further attempts could expand to various formats.

**Author Keywords**

Automated News; Automated Journalism; Computational Journalism; Natural Language Generation

**ACM Classification Keywords**

K.4.3 [Computers and Society]: Organizational Impacts; J.4 [Social and Behavioral Science]: Sociology; H.3.5 [Online Information Services]: Web-based services

---

Permission to make digital or hard copies of part or all 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. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

TVX'16, June 22-24, 2016, Chicago, IL, USA

ACM 978-1-4503-4067-0/16/06.

<http://dx.doi.org/10.1145/2932206.2933561>

## Introduction

In January 2015, an article titled "Apple Tops Street 1Q Forecast" was released <sup>1</sup>. It was a piece of ordinary finance news published by the Associated Press (AP), written by an algorithm developed by Automated Insights <sup>2</sup>. Subsequently, the convergence of journalism and computational science led to the advent of the emerging field of "computational journalism." Computational journalism can be defined as an area where "computation advances journalism by drawing on innovations in topic detection, video analysis, personalization, aggregation, visualization, and sensemaking" when stories including news and reports are "discovered, presented, aggregated, monetized and archived [3]." In the realm of computational journalism, "automatic journalism," also referred to as "algorithmic journalism" or "robot journalism [6]," specifically refers to the field where algorithms are used to automatically generate news stories [2].

Automated journalism is often applied in the areas of weather [8], finance <sup>3</sup>, and sports [1], where the underlying data are accurate and reliable. This is because automated journalism requires high-quality data that is structured and accurate [2]. Despite the growing number of applications of automated narratives in various domains, no attempts have been made that applied to the television industry.

This paper suggests a methodology of automatically generated news about TV ratings to increase efficiency by automating repetitive tasks. The framework goes through the processes of data gathering, identifying key events by a predefined algorithm, generating a story, and publishing.

<sup>1</sup><http://finance.yahoo.com/news/apple-tops-street-1q-forecasts-213944804.html>

<sup>2</sup><https://automatedinsights.com/>

<sup>3</sup><https://blog.ap.org/announcements/automated-earnings-stories-multiply>, <http://www.forbes.com/sites/narrativescience>

The algorithm that determines the structure of the story is defined by analyzing existing news of TV ratings reflecting key variables. Furthermore, not only TV ratings but also diverse data such as online streaming viewers, social opinions, and online clip views could be the topic of automated news in the television industry.

## Related Work

### *Automated journalism*

Automated journalism is the process of automatically generating stories with predefined algorithms [2]. An algorithm refers to a "well-defined computational procedure that takes some value, or set of values as input and produces some value, or set of values as output [5]." Algorithms include the process of input, throughput (processing), and output. Therefore, algorithms are designed to logically process the data for a meaningful output. This also applies to algorithm-based automated journalism. To produce news stories automatically, we need high-quality data (input) and reliably designed algorithms (throughput) to create relevant news stories in narrative form (output). The typical companies providing automated journalism solutions are Automated Insights <sup>4</sup> and Narrative Science <sup>5</sup>. Their solutions are not limited to news or journalism areas. They also cover financial services, business reports <sup>6</sup>, etc. However, far too little attention has been paid to the television industry.

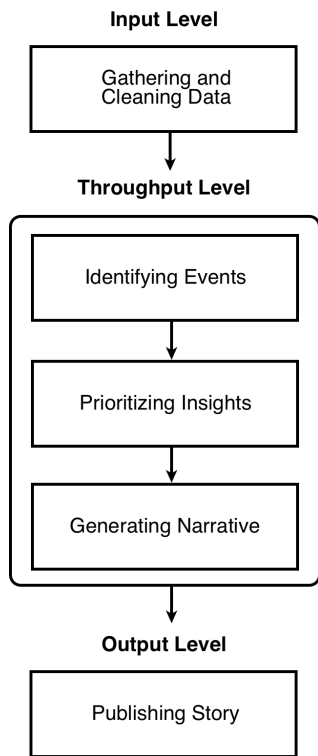
### *TV ratings*

In the television industry, one of the most important indexes is TV ratings, which determine the popularity of TV programs and charges for advertisement. Among TV ratings, the demographic of viewers in the age range of 18 to 49 is crucial because of their buying habits [9]. Representative

<sup>4</sup><https://automatedinsights.com/>

<sup>5</sup><https://www.narrativescience.com/>

<sup>6</sup> <https://automatedinsights.com/solutions/>



**Figure 1:** Framework of Automated Journalism

TV ratings include the ratings points, share, and total viewers among adults 18-49 <sup>7</sup>.

Even though ratings points/share is the major evaluation criterion for the value of TV programs, the diffusion of digital media has weakened its impact. Because of the widespread use of digital media, the watching pattern of TV programs has changed. According to Ericsson [4], 86 percent of smartphone users watch video content on their phones. Applying the new trend of media use, "social rating" <sup>8</sup> is suggested as the new measure of audience engagement. It measures both activities (authors, tweets) and reach (unique audience, impressions) on Twitter. The research restricted the object of analysis to traditional TV ratings, after which the social data could be included as part of the input data if the data processing method was well defined.

## System Overview

### Framework

Primarily, natural language generation is explained within the framework of algorithmic selection along the Input-Transform-Outcome (ITO) model [6]. Based on the natural language generation process [7], the framework of automated news generation of TV program ratings has been subdivided into five stages: 1) collecting data, 2) identifying interesting events, 3) prioritizing insights, 4) generating the narrative, and 5) publishing the story (Figure 1).

- Collecting data: The first step is data collecting and cleaning. Data could be obtained from databases, files, logs, system interfaces, etc. The original source of the raw data requires a processing procedure,

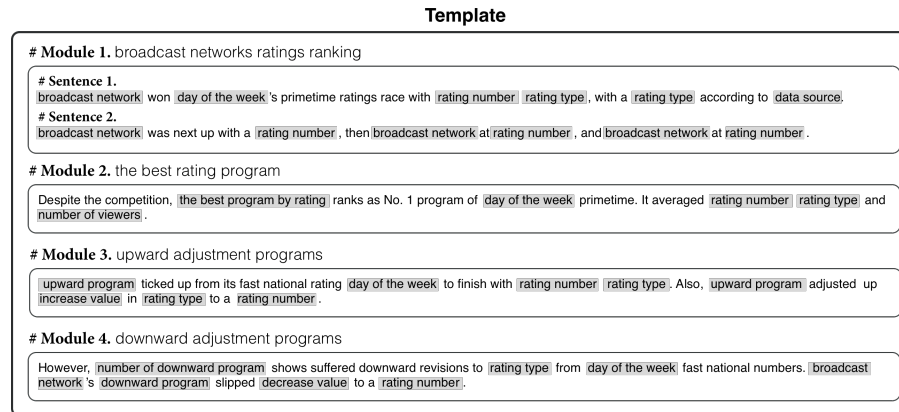
<sup>7</sup>Rating: Estimated percentage of the universe of TV households (or other specified group) tuned to a program in the average minute. Ratings are expressed as a percent. Share is the percentage of television sets in use that are tuned to the program.

<sup>8</sup><http://www.nielsensocial.com/>

which includes merging, subsetting, and transforming to be ready for the next level. After gathering data, a procedure for identifying useful information (structured data) is needed for the intended output. Data is accessible through APIs or databases.

- Identifying interesting events: This is the stage that identifies the input-output relationship with a specific algorithm. Predetermined rules are applied to the algorithm to choose and process data for meaningful outcomes.
- Prioritizing insights: In this stage, meaningful variables and events are selected. For Nielsen television ratings, "ratings points/share" and "total viewers" could be the variables. In this case, fluctuation (upward/downward) and ranking (of broadcast networks or programs) of those indexes are important events. In addition to traditional ratings, with the widespread use of digital media, social media indexes such as "Nielsen Twitter TV Ratings" <sup>9</sup>, a new method to assess TV programs' influence, could be included as a significant factor. After important events in the data are defined, the algorithm categorizes and identifies those events by importance. This is the process of creating the mood of the stories, which would be specified afterward.
- Generating the narrative: Based on the algorithm, the news narrative is generated by predefined structures and specific sentences.
- Publishing the story: Finally, the completed story is uploaded and published. In this stage, the story could be uploaded either automatically or after examination.

<sup>9</sup><http://www.nielsensocial.com/product/nielsen-twitter-tv-ratings/>



**Figure 2:** Example of News Article Structure

### News Article Structure

The structure of news has several variables, defined below (Figure 2). Figure 2 illustrates the relationship between four variables.

- **Template:** A template is the whole structure of the news story. Based on selected events through the process of prioritizing insights, the algorithm chooses one from the predefined set of templates. In this example, only one template—one type of news article—is suggested. Nevertheless, it would expand to various templates, each of them containing a different point of view.
- **Module:** A module is the building block of the template. Each of the modules contains critical issues, such as the daily ranking of broadcast networks, the best rating program, etc. Some modules could be ruled out if they are not so significant to the situation.

A module is composed of sentences that are predefined by the developer.

- **Sentence:** Sentences compose the module. It is the unit of the module that actualizes the main topics into natural language.
- **Data:** Data are the initial input of the overall process of automated journalism. By assembling, cleaning, and analyzing data, the whole narrative is completed.

### Discussion & Conclusion

We have suggested a way of automating the generation process of news in the TV industry. By automating processes such as data gathering, data analysis, writing stories and reports, and publishing, the speed and accuracy of the working process could be increased [2]. The suggested framework and news article structure would be the foundation for further research on TV news generation in the future.

There are several points that might be improved in future research. Above all, by implementing a working prototype, more experiments that could evaluate the results may be developed in order to identify the different qualities of the articles generated by the algorithm and those written by human journalists.

In addition, the type of news article template could be further expanded by analyzing significant data in the TV industry. Not only TV ratings but also public opinions about TV programs could serve as a form of report by analyzing social data. More comprehensive research on news format and data processing may be used to develop automated journalism in the TV industry.

### Acknowledgments

This work was supported by the CPRC program of MSIP/IITP. (IITP-2015-H8201-15-1004)

This study was financially supported in part by the Institute of Communication Research, Seoul National University.

### REFERENCES

1. Nicholas D Allen, John R Templon, Patrick Summerhays McNally, Larry Birnbaum, and Kristian J Hammond. 2010. StatsMonkey: A Data-Driven Sports Narrative Writer.. In *AAAI Fall Symposium: Computational Models of Narrative*.
2. Tow Center. 2016. Guide to Automated Journalism. (7 January 2016). Retrieved from <http://towcenter.org/research/guide-to-automated-journalism/>.
3. Sarah Cohen, James T Hamilton, and Fred Turner. 2011. Computational journalism. *Commun. ACM* 54, 10 (2011), 66–71.
4. Ericsson ConsumerLab. 2015. TV and Media, 2015. (September 2015). Retrieved from <http://www.ericsson.com/res/docs/2015/consumerlab/ericsson-consumerlab-tv-media-2015.pdf>.
5. Thomas H Cormen. 2009. *Introduction to algorithms*. MIT press.
6. Konstantin Nicholas Dörr. 2015. Mapping the field of Algorithmic Journalism. *Digital Journalism* (2015), 1–23.
7. Ehud Reiter, Robert Dale, and Zhiwei Feng. 2000. *Building natural language generation systems*. Vol. 33. MIT Press.
8. Ehud Reiter, Somayajulu Sripada, Jim Hunter, Jin Yu, and Ian Davy. 2005. Choosing words in computer-generated weather forecasts. *Artificial Intelligence* 167, 1 (2005), 137–169.
9. Michael Storey. 2009. The TV column: Not in 18–49 age group? TV execs write you off. *Arkansas Online* 23 (2009).