Publishing semi-automated content at the BBC

Experiments in local stories at a national outlet

Conor Molumby
BBC News Labs
London, U.K.
conor.molumby@bbc.co.uk

ABSTRACT

The challenges involved in publishing semi-automated local content at a large national outlet will be identified, and possible solutions will be discussed. Issues that do not arise in the publishing pipeline when individual stories are written by a journalist but that now arise when using some automation have been linked to their location in the process at the BBC.

Three steps in the process have been identified as requiring significant adjustments to the typical story publishing methodology at the BBC: **Crafting** the story content, the **editorial processing** of the draft story, and the **distribution channel** of completed stories.

Three projects in which semi-automation was utilized will be discussed as case studies to illustrate the issues and the solutions attempted.

The examples are: 300 variations of a story based on tree planting statistics, approximately 7,000 variations of a story on the changing high street, and a general election results story for all 650 constituencies in the UK.

Crafting stories using semi-automation requires a new form of input from journalists. Whilst many examples illustrating the use of automation exist in business journalism, the BBC's experiments have focused on other types of stories that use data. This requires a journalist to think in an abstract way about the different scenarios that may arise in the data and build narrative structures around them in a natural language generation (NLG) template.

The level of abstraction that is required for a journalist to think about story building as a craft that accounts for all scenarios as opposed to writing one bespoke piece of content based on a single occurrence is something that is tackled differently depending on the project. For local tree planting stories, the journalist already had the dataset and was able to examine the extremities of it while building their template.

In the case of preparing a general election results story the process was very different. The pressure of a live, high-profile news event meant that the concept of the template had to be explained to seasoned political journalists and editors. Several iterations of it had to go through many levels of editorial approval at the BBC before it could be used on election night. Showing how each scenario was crafted by journalists - writing sentence structures in English or Welsh, and coding logic in the tool's markup language - helped to build trust.

It was found that the best way to communicate the accuracy of the stories to senior editorial stakeholders was to create examples based on template output in different scenarios. Seeing old election data processed through the template also assured them that the system would reliably produce the correct narrative based on the live data.

The **editorial processing** of a draft semi-automated story bears some resemblance to the checks and balances a one-off story typically goes through at the BBC. Editors sub copy and check data and facts. The fundamental difference is that all of this output has come from a system and not a human journalist. When general election stories were sub-edited by a newsroom team the machine built up trust with them in the same way an editor grows to depend on a reliable reporter over time. Editors' trust grew in response to the number of stories they read. Editors did need to adapt to how to query the creator of the story. Instead of asking a journalist how they reached a particular conclusion the editors referred to the source data.

Systems and processes had to be built to notify editors when stories were generated and to allow them to keep track of the high volume of stories created using automation. Keeping a human sub-editor in the loop has helped to accelerate the adoption of the tool at the BBC.

The **distribution channel** used for semi-automated content must have a granularity that matches that of the data used as the determinant of the number of stories. This proved to be an issue with the 7,000 high street stories. The BBC does not currently have a way to exclude stories from a user's local news feed. This means users could end up seeing many versions of the same semi-automated piece of content if they set a wide geography for their local news. A lookup tool was built for end users for these stories.

Election stories were efficiently published because the BBC website includes a specific page for each constituency in the U.K. Surfacing these 690 stories to the correct webpage proved easier than publishing the tree planting stories because of the direct matching of stories to pages inherent in the former project.