Digital Creativity Support for Original Journalism

Neil Maiden, Konstantinos Zachos, Amanda Brown, Dimitris Apostolou, Balder Holm, Lars Nyre, Aleksander Tonheim, Arend van den Beld

Abstract

The decline in circulations and revenues resulting from the digitalization of news production and consumption has led to a crisis in journalism. Journalists have less time to research, investigate and write original stories, leading to problems for our democratic processes and holding the powerful to account. This paper reports the architecture, features and rationale for new digital creativity support designed to support journalists to discover more original angles on stories. It also summarises the evaluation of the tool’s use in 3 newsrooms.

Original journalism

Journalism involves the search for and critical analysis of information (McNair 1998). How journalists discover and select sources of this information is important, to avoid bias, to be credible and trusted, and to create angles with which to generate new stories of value to readers.

Journalist creative thinking, to discover and generate new associations during this search and analysis of information, contributes to the generation of new stories. Journalists are known to seek opportunities to develop new creative skills with which to discover information (Malmelin & Virta 2016). Applying these skills enables journalists to maintain control over their work (Witschge & Nygren 2009). And emerging forms of investigative journalism demand new creative search and association skills (Gynnild 2014).

However, discovering and examining information sources about complex stories takes time – time that journalists increasingly lack as news organizations reduce staff numbers (Sjøvaag 2014). The digitalization of news production and consumption has led many news businesses to become uncompetitive. Some work practices are slow to change due to conflicts with journalist professional values for autonomy (Ekdale et al. 2015). Therefore, as coping strategies, journalists often use subsets of available and familiar information sources to create stories, which in turn can reduce the diversity of angles used to report stories.

Although journalism is one of the creative industries, explicit support for the creative skills of journalists is rare. For example, it is not one of the 5 journalist capabilities reported in (Cohen et al. 2011), and few digital tools support journalist creativity.
INJECT was a new digital tool that was designed to support journalists to discover new associations with which to generate stories with angles more novel and valuable than stories published previously. It integrated creative search algorithms with which to discover information in published news stories and interactive support to form new associations with this information during creative thinking. It was designed to contribute to journalist engagement in professional-level creative work, i.e. work that generated income and provided them with a living (Kauffman & Beghetto 2009). This work included writing stories that were creative, i.e. judged to be novel and valuable (Maher & Fisher 2011) through the application of the creativity skills (Amabile & Pratt 2016) of the journalists.

Existing creativity support technologies for journalists

Digital tools that enhance the creativity skills of journalists are rare. One exception is the Story Discovery Engine, which used artificial intelligence algorithms for investigative reporting (Broussard 2015). The Tell Me More system mined the web for similar stories reported by different sources and extracted text that offered new information in the form of quotes, actors and figures (Iacobelli et al. 2015). Both of these tools had similar objectives to INJECT, but were not framed as creativity support tools for journalists. The SocialSensor news app surfaced fast moving trends from social media content, but revealed biases arising from such content (Thurman 2016). Many different data visualization tools support journalists to make sense of, for example, social media content (Diakapolous et al. 2010). However, none supported human creative thinking to discover new angles on news stories.

To workaround the lack of bespoke tools for information search and analysis tasks, many journalists use generic search tools such as Google (Machill & Beiler 2009), but these lack explicit support for human creative thinking about news stories.

Unlike in journalism, digital creativity support has been implemented for professionals in other creative industries such as the performing arts, music, and film and television. Examples of the digital support include StoryCrate, a collaborative editing tool developed to drive users’ creative workflows within a location-based television production environment (Bartingdale et al. 2013) and Trigger Shift, which appropriated information technologies into performance art in theatre (Schofield et al. 2015). Other domains for which digital creativity support has been developed include theatre, scientific discovery, and caring for older people.

Therefore, to fill the gap in journalism, new digital creativity support was developed that aligned with the work practices, tools and values of journalists. The resulting tool was called INJECT.

Designing INJECT with journalists
INJECT’s design was informed by established cognitive models of creative thinking. Most of these models describe dual processes of developing and evaluating ideas to generate outcomes that are both novel and valuable (Kaufman & Beghetto 2009, Maher & Fisher 2011). Developing ideas is a divergent and associative process that can be spontaneous and deliberate, and involves retrieving relevant items from memory and generating associations with new information (Finke et al. 1992). By contrast, evaluating ideas is more analytic, but can be interleaved tightly with developing ideas (DiPaolo & Gabora 2007).

Therefore, INJECT was designed to support journalists to discover new information, generate associations between this information and items from memory to discover new angles on news stories, and evaluate these angles quickly during story development.

To align INJECT to these work practices tools and values, journalists were included in the tool’s design. Interviews were held with experienced and inexperienced journalists to discover problems, requirements and constraints. Paper-based then digital wireframes of the INJECT tool were developed and presented to professional journalists. New releases of the working INJECT software were prototyped for their usability and impact with professional and student journalists.

The user-centred design process uncovered 3 important values that most journalists held about their work – values that the INJECT tool was designed to uphold.

The first value was the importance of discovering information already reported in verified newspapers, as opposed to in unverified sources, as the starting for discovering new angles on stories. Even though it was argued that published news might constrain their creative thinking, most journalists expressed a preference for it to direct the discovery of new associations and angles. Feedback on prototypes revealed 3 specific types of verified news information were effective for discovering angles: 1) published news stories similar but not the same as the new story being written; 2) entities such as people, places and organizations that might relate to these new stories, and; 3) guidance for directed creative thinking to develop the stories. INJECT was designed to direct journalists to generate new associations between information discovered in similar stories and in entities referenced in these stories.

The second value was to recognise the existing creativity skills of journalists. Many who engaged in INJECT’s design initially rejected the need for digital support for their creative thinking. After all, journalism is one of the creative industries, and many chose it as a profession to be creative. Instead, journalists identified with the need to generate new angles more quickly.

The third value was that creative thinking was not separate from but part of everyday journalist work. Instead, journalists sought support for more original journalism that was embedded in daily work tasks and tools, such as text editors.
The INJECT tool

The INJECT tool was implemented with natural language processing, multi-language creative search, and interactive creativity support capabilities. To provide journalists with a sufficiently large external information source to discover associations with, it indexed content from millions of verified stories published by hundreds of news titles in multiple languages.

The INJECT tool’s three-tier architecture is shown in Figure 1. The interaction layer was a sidebar designed to be simple, fit with existing work practices, and encourage journalists to discover new angles on stories quickly without learning new skills.

![INJECT tool diagram](image)

**Figure 1. The INJECT tool’s three-tier architecture, showing its layers, services and external information sources**

The application layer was composed of services designed to generate large numbers of possible associations between information that journalists were writing about using indexed news content from millions of already published news stories.

These services retrieved this content from INJECT’s data layer, called the creative news index, which was designed so that the discoverer service could undertake divergent creative searches that were more sophisticated than were possible with existing web search and news site APIs. The index was populated by the presser service, which indexed millions of verified news stories as possible starting points for discovering new angles on stories. The text processor service was invoked by the presser to make sense of and generated indexed content from published news, and by the discoverer to expand creative search queries.

**INJECT’s presser**
The presser generated indexes of millions of verified news stories that could be retrieved, on request, as starting points for journalists to discover associations with which to generate new angles on stories. It had a crawler component that fetched news stories to index from open RSS feeds, and an importer component that fetched stories from accessible newspapers’ archives.

The crawler was directed to fetch verified news stories from 1105 predefined RSS feeds published by 380 diverse news titles in 6 languages. These feeds, titles and languages were selected by INJECT’s editorial team to generate indexes of diverse views and angles on news, and ranged from major daily newspapers in the United States, regional newspapers in the Netherlands, and tabloid titles in the United Kingdom. On a normal news day, it fetched about 15,000 verified stories. Stories from high-frequency feeds were fetched every 30 minutes, others every 12 hours. During each fetch cycle, the crawler automatically read all news stories accessible via the URLs in the each RSS feed, removed navigation links, adverts and embedded media such as links, images and videos, and sent the remaining text string, along with story’s author, URL, image URL and published date, to the text processor service. This text string, author, date and URLs provided a rich external information source with which journalists could discover and generate new associations and angles on news stories.

The importer component was similar to the crawler, but was directed to fetch stories from local JSON files. It was developed to address the need of news organisations to use their own stories as starting points for more original journalism. Like the crawler, it also sent a text string, along with author, URL, image URL and published date, to the text processor service.

**INJECT’s text processor**

The text processor service generated new entries to add to the creative news index by analysing the natural language text string of each fetched news story with:

- Named entity extraction mechanisms to index stories using real names such as people and places. The mechanisms that treated candidate named entities as groups of consecutive words describing a concept such as a person (e.g. Tawakkol Karman), location (e.g. Sana’a), organisation (e.g. United Nations) or object (e.g. war crime). This enabled the processor to extract entities with which journalists might discover associations not described in the text, for example the entity Sana’a from the text the capital of Yemen. After experimentation with alternatives, the processor invoked the DBpedia Spotlight (2017) and Polyglot (2017) services. Spotlight annotated mentions of DBpedia resources using entity detection and disambiguation algorithms with adjustable precision and recall, which were used to refine INJECT’s sensitivity to news content using measures such as entity prominence, topical pertinence and disambiguation confidence. Polyglot implemented named entity extraction, speech tagging, sentiment analysis, morphological analysis, and transliteration in all of INJECT’s 6 target languages – English, German, Dutch, French, Italian and Norwegian. It
could detect, for example, that *Forente Nasjoner* is Norwegian for the entity *United Nations*, the international organisation founded in 1945;

– Automatic parser mechanisms that detected nouns and verbs to index stories using common objects and actions. The parsers split news text into sentences then applied part-of-speech tagging to mark up words as belonging to lexical, part-of-speech categories. Shallow parsing was applied to generate a machine understanding of the structure of a sentence without parsing it fully into a parsed tree form. The output was a division of the text’s sentences into a series of words that, together, constituted a grammatical unit. To select candidate objects and actions from these units with which journalists might also discover associations, the mechanism applied lexical extraction heuristics on a syntax structure rule-tagged sentence. For example, the processor parsed the news headline *The Yemen war in the world’s worst humanitarian crisis* to extract the nouns such as *war, world and crisis.*

**INJECT’s creative news index**

For each fetched story, the creative news index generated a new entry composed of all extracted named entities, objects and actions and frequencies of occurrence, the author, URL, image URL and publication date. A typical entry for a news story of 400 words was composed of between 30 and 50 entities, objects and actions. Early prototyping of the INJECT tool revealed that indexes with this volume and type of content were sufficient to generate new associations that journalists reported could be effective for discovering new angles.

All index entries were uploaded to an external Elasticsearch cluster to be manipulated by the discoverer’s creative search algorithms. Elasticsearch is a scalable open source search engine with a REST API that provides scalable, near real-time search. This performance was essential to support journalists to discover new angles on stories more quickly. In July 2019, the Elasticsearch cluster held over 14 million entries, with another 450,000 new entries being added each month.

**INJECT’s sidebar and discoverer**

Journalists interacted with the INJECT sidebar to discover new associations and angles on stories. The sidebar was designed to provide journalists with index information and features and generate new associations with this information without opening another application. To work within the space constraints of the widget, the sidebar was implemented with mouse hover-boxes and information that journalists could use to discover associations quickly. Its design also supported journalists to flip quickly between ideation and evaluation processes during story development.

Figure 2 depicts use of the sidebar in the Google Docs editor to discover associations leading to angles for a new story about the Yemen humanitarian crisis. The sidebar was also implemented for the *Wordpress, Adobe InCopy* text editors, *Google Chrome*
web browser and content management systems that use the TinyMCE text editor, as well as a separate web application that a journalist could reshape as the sidebar.

Figure 2. The INJECT sidebar on the right side of the Google Docs text editor. A journalist writing a new story about the Yemen crisis is presented with news articles and information about places, things, people and organisations with which to discover associations leading to new angles.

If the journalist highlighted text in the editor, the sidebar invoked the text processor service to extract named entities (e.g. Yemen), nouns and verbs (e.g. crisis, ecological) as candidate topics to present at the top of the sidebar, see Figure 2. This feature was implemented to increase the sidebar’s usability and enabled journalists to work more quickly.

The journalist could then use the icons beneath these topics to select between 6 predefined creative strategies that mimicked the strategies of experienced journalists (Maiden et al. 2018). These strategies were implemented in the discoverer service to retrieve creative news index entries with:
A. Quantified information associated with the topics;
B. Information about people associated with the topics;
C. Information about events associated with the background of the topics;
D. Information about future consequences associated with the topics;
E. Data sets and visualizations associated with the topics;
F. Comical information associated with the topics.

For strategies A-E, the discoverer:

1. Disambiguated each noun topic term by discovering its correct sense in the online lexicon at WordNet using context knowledge from other terms in the query (e.g. that crisis is an unstable situation of extreme danger or difficulty rather than a crucial stage or turning point in the course of something). It then expanded each term with other terms with similar meanings (e.g. the term crisis with the above sense is synonymous with exigency and flashpoint) and included these terms in the search query. Term sense disambiguation and query expansion was implemented to retrieve index entries that were different lexically but related semantically to the topic terms, so that journalists could generate new associations based on different types of semantic similarity;

2. Invoked an Elasticsearch search via the news API with the expanded query terms and logic operators set by the journalist to control search breadth. Elasticsearch returned a set of indexed entries that achieved a threshold match score in response times acceptable to journalists;

3. Scored the returned index entries for relevance based on the frequencies of original and expanded query terms in the title of each story, to prioritise entries with headlines related to topic terms. This scoring mechanism was implemented to reflect the structure of most news stories with the most important information at the start of stories;

4. Filtered the scored index entries using constraints specified for the selected strategy, so that journalists were presented with information to form associations consistent with that strategy. For example for quantified information (A), it filtered to retain entries with a minimum threshold of 100s of quantity, measure and value keywords, for example Sterling, population and actual numbers. For information about events associated with the background of the topic terms (C), it filtered to retain entries with more than 500 words of content and a minimum threshold of 100s of keywords indicative of background articles such as cause, impact and studies from sources such as the Economist and the New York Times. And for information about people (B), it generated orders of entries that reference a person entity named in a minimum number of entries.

The discoverer sent JSON representations of each remaining index entry to the sidebar to display as a news card. By contrast, for strategy F, discoverer generated simpler keyword queries that searched the caption text of over 60,000 political cartoons accessed by INJECT via an API from an external database.

This automation of information discovery was designed to enable journalists to commit more cognitive resources to generating associating and evaluating ideas.
The sidebar presented the retrieved information as a scrollable sequence of news cards. Journalists could select the information to view using sidebar features to sequence the news cards by relevance, date of publication or random, and to present news published within selected periods.

Each news card in the sidebar presented the title, publication, date, first sentence, and 10 randomly selected entities. Clicking on the title opened the original new story or cartoon, at source, in a new browser tab. Positioning the cursor over each rectangle presented a pop-up creativity spark generated for that places, things, people and organisations. This feature was implemented as a mouse hover-over to enable journalists to explore multiple sparks and discover different associations quickly. The sparks themselves were designed to direct the deliberate generation of associations and ideas by journalists. Each was generated by the sidebar from a predefined set of spark types to direct journalists to think about, for example, the history and relevance of places, the motives of people and their opponents, the future and emotional impact of objects, and available data about organisations.

Figure 4 shows these features in 3 different INJECT sidebars presented for different angles using the same information about the Yemen humanitarian crisis.
Figure 3. Three INJECT sidebars presented for the new story about the Yemen crisis, showing (from left-to-right) information to support a background angle, a people angle, and a comical angle based on political cartoons.
Figure 4. Three INJECT sidebars presented for the same new story about the Yemen crisis in Norwegian, showing (from left-to-right) showing use of a background information angle, a people angle, and use of creativity sparks in list form.

The sidebar also presented other styles of news card showing only entities, word clouds, and sparks in list form. These other styles, shown in Figure 5, were added to reduce comparisons with Google search that reduced journalists’ expectations for creativity support.

Furthermore, to support journalists to evaluate as well as discover ideas, the sidebar launched Google web searches in new browser tabs from within INJECT, to retrieve information with which to analyse and critique ideas, see Figure 5.
The INJECT tool was tested by journalists working in multiple languages. When sufficiently robust, it was evaluated in different newsrooms.

**Evaluating the INJECT tool in 3 newsrooms**

The INJECT tool was installed in the newsrooms of 3 regional newspapers in Norway to investigate the effectiveness of its creativity support. One research question explored was whether journalists produced news stories that were more novel and valuable with INJECT’s support.

INJECT was introduced into the daily work of 4 journalists in each of the 3 newspapers for 2 months in 2018, for use in Norwegian and English. The 12 journalists received INJECT training and helpdesk support, and were encouraged by their editors to use INJECT. During the evaluation, the numbers of English-language entries in the creative news index increased from 2.7 million to 3.2 million and Norwegian-language entries from 260,000 to 300,000. The index also included 62,160 Norwegian-language articles from archives of the 3 newspapers generated by the importer component, and INJECT also searched over 50,000 digital cartoons. The journalists used INJECT’s web application version.

To investigate the research question, news stories produced by the journalists with and without the support of INJECT were rated by 7 individuals with journalism expertise and/or knowledge of the regions of the 3 newspapers, see Sidebar A.

Three of the 7 judges were domain experts in journalism – associate professors of journalism at local higher education institutions. The other 4 had roles that equipped them with extensive local knowledge, as head of information at a regional institute in business and trade, two local business leaders in tourism, and a retired legal stenographer. All 7 lived in the regions covered by these newspapers.

Each judge was assumed to be able to rate 40 news stories accurately in the available 3-hour period, so each rated 20 news stories that journalists had written...
with support from INJECT and 20 written without it in the same period 12 months earlier. A random number generator algorithm at random.org was used to select the 40 news stories, and numbers of stories proportionate with the total number of stories written by each journalist with the support of INJECT were selected. The 40 news stories were then randomly ordered in a questionnaire using another algorithm at random.org, anonymized and presented with two 1-7 scales to capture each judge’s novelty rating and value rating of each news story.

Sidebar A. The expert judgement process used to rate news stories

INJECT was used in all 3 newsrooms. No major technical problems were reported. A total of 72 published stories were written with the support of INJECT by 10 of the journalists. Journalists used already-published news stories as effective starting points for new angles on stories. Based on the expert analysis, a Mann-Whitney test revealed that the novelty ratings were greater for the news stories written with the support of INJECT (Mdn=3) than without the support of INJECT (Mdn=2), U=6997.5, p<0.0001. INJECT use was associated with an increase on the novelty of news stories, albeit from ratings that indicated low novelty of most non-INJECT news stories.

In contrast, a second Mann-Whitney test revealed that the value ratings were not greater for the news stories written with the support of INJECT (Mdn=5) than without the support of INJECT (Mdn=5), U=9156, p>0.05. The average value rating of all of the news stories was 4.7 out of 7, and the lowest and highest average value-rated articles were 3.71 and 5.86. This was unsurprising, given that all of the news stories had passed through editorial processes.

Most of the journalists needed time to learn to use INJECT, and many reported comparisons to Google: “You need to adjust slightly, because we are used to search engines that give us the most popular hits”. INJECT use was related to journalist attitudes. Four younger journalists in one newspaper who were open to new technologies and worked more autonomously used INJECT more frequently. By contrast, more experienced journalists were less willing to adopt INJECT after the evaluation: "We seem to have certain stubbornness against using INJECT and other tools like it".

The evaluation in the 3 newsrooms revealed that the journalists did produce news stories that were more novel if not more valuable with support from INJECT. However, all were still published, indicating sufficient value for purpose. One interpretation of this result was that the stories written without the tool’s support had value but lower novelty, i.e. the stories were not creative. Articles written with the tool’s support had increased novelty but not increased value. In a strict sense, these articles were more novel rather than creative, but still had sufficient value to publish.

More results are reported in Maiden et al. (2019).
Conclusions

Demonstrations of INJECT to other news organisations reinforced our judgment that digital support for journalist creative thinking is rare (Broussard 2015, Iacobelli et al. 2015). However, positive receptions revealed the potential of INJECT to support journalist creative thinking.

To uphold the 3 journalist values uncovered during design, INJECT’s interactive support was separated from indexing published news. The sidebar design enabled journalists to access INJECT’s guidance in as few as two clicks, without leaving the text editor. It demonstrated how to establish digital support for creative thinking as part of journalists’ daily work tools, although more evaluations are needed.

INJECT’s creative news index is now an important asset of more than 500million pieces of news information for computational manipulation. New computational analyses under development will detect patterns, biases and angles on news shown to be novel, and hence creative. One will analyse differences in topic reporting in different languages to generate angles underreported in a target language. Rolling out new INJECT versions with these features will support news businesses to remain competitive and fulfil their role in liberal democracies.

Acknowledgements

The research reported in this paper was supported by the EU-funded H2020 723328 INJECT innovation action.

References


