
This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: http://openaccess.city.ac.uk/20291/

Link to published version:

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.
Creative Information Exploration in Journalism

Konstantinos Zachos1, Dimitris Apostolou2,3, Fotis Paraskevopoulos3, Stergios Ientsek3, Neil Maiden1, Gregoris Mentzas3

1CASS Business School, City University, London, UK
2Department of Informatics, University of Piraeus, Piraeus, Greece
3Information Management Unit (IMU), Institute of Communication and Computer Systems (ICCS), National Technical University of Athens (NTUA), Athens, Greece

Abstract: Existing digital tools used by journalists such as content management systems and search engines focus on helping journalists find relevant information and organize their creative work; they do not provide support for discovering creative angles to investigate. To meet this need in journalism, we have researched new creative search algorithms that manipulate the journalist’s primary medium – written information – and developed an innovative creativity and productivity support tool deliberately for use by journalists. In this paper, we analyse the creative tasks journalists perform when developing new stories, we provide an overview of requirements, which have been addressed by the INJECT digital creativity support tool for journalists. Further, we explain the intended use of the system using a walkthrough scenario and present our early experiences from deploying the tool in journalist’s working environments.

Keywords: creativity, journalism, creative search

I. INTRODUCTION

The last years have seen the emergence of new ICT solutions to support and improve the creativity of workers in different professions. New digital creativity support tools have been developed for and evaluated successfully in sectors that range from manufacturing and design [32] to the biological sciences and the care of older people with dementia (e.g. [31], [33]). New forms of digital support have also been developed to enhance worker productivity in targeted creative industries such as the performing arts (e.g. [23]), music, film and television (e.g. [3]). Indeed, recent studies have revealed critical sweet spots in the use of digital creativity support on production lines that increases the productivity as well as creativity of workers.

In contrast, there has been little systematic deployment of digital support for creativity and productivity in journalism – even though it is one of the creative industries, recognised to be an important cornerstone of functioning democracies [2]. Existing digital tools used by journalists such as content management systems and search engines focus on helping journalists find relevant information and organize their creative work [15]; they do not provide support for discovering creative angles to investigate. Interactive creativity support tools exist for different creative industries (e.g. [12]; [1]), but newsrooms lack the digital support to discover possible creative angles on news stories, i.e. stories that are novel and have value to their readers [35].

To meet this need in journalism, we have researched new creative search algorithms that manipulate the journalist’s primary medium – written information – and developed an innovative creativity and productivity support software tool deliberately for use by journalists. In this paper, we analyse the creative tasks journalists perform when developing new stories, we provide an overview of requirements, which have been addressed by the INJECT digital creativity support tool for journalists. Further, we explain the intended use of the system using a walkthrough scenario and present our early experiences from deploying the tool in journalist’s working environments.

II. JOURNALISM CREATIVE TASKS AND REQUIREMENTS

A. Creativity in Journalism

An analysis of the literature on journalism provided mixed messages concerning creativity in journalism [9]. The genres of print journalism such as feature writing, opinion pieces and literary journalism are considered widely to provide an outlet for creativity [8] [13], [17], [19], [20] [24]. On the other hand, the news genre is regarded as formulaic and restrictive and is thus seen as not being creative [20]. A similar controversy seems to appear in research by scholars, some of them do call for journalists to be creative [7] [6] [29], [30], while other texts argue that journalism’s structures constrain journalists [10], [11] [27]. Arguably, the aforementioned controversies may be attributed to the different definitions of creativity that scholars follow. The so called “romantic” view of creativity calls for individuals to be free from constraints and structures to be able to create and for a creative product to be completely original and come from nowhere but the imagination of the individual [9]. On the contrary, research into creativity has clearly shown that, in line with a rationalist view, creativity is always embedded in previous works, it is always the product of a system, rather than solely attributable to an individual, and there are always structures to constrain and enable an individual in their creative process [9]. Our work follows the later definition and aims to leverage creative outcomes within typical activities and tasks that journalists undertake.

B. Indicative Scenario

We used scenarios to both discover and analyse requirements emerging from the application partners. Each scenario is used to define discrete, contextualised actions that both permitted not only the analysis of requirements associated with these actions, but also the human interpretation of the scenarios to discover and validate requirements. The remaining of this section outlines an indicative scenario and discusses emerging requirements from a system that would facilitate the creative outcomes of journalists.
Hallingdølen is a Norwegian language local newspaper published in Ål, Norway and serves Hallingdal. Hallingdølen was established in 1936 and founded by community members. The paper is published three times per week and has its headquarters in Ål. It was published in broadsheet format until 1995 when it switched to tabloid format. To illustrate the story development process, let us focus on a storyboard about writing a story about the reduced local post service affecting the local community. Hence, easy access and search of the newspaper archives is essential before even starting the investigation.

Once key story components have been identified, a common activity that the journalist pursues is to discover and interview specific people. In investigating the impact of the reduced post service, the journalists need to find facts such as the number of households affected. The journalists also need to find background information that will help her understand the global picture. E.g., How many are dependent on receiving mail every day? Have similar things happened in other countries? What kind of ideology/politics lie behind these reductions in service? etc. In the end, journalists need to put all information together, report the facts and discuss them in the article.

A main challenge in the aforementioned scenario that INJECT is expected to address is the amount of time needed to write a story, from the ideation phase to the published article. In particular, INJECT can accelerate the time needed to investigate the background of a story, without having to spend too much time finding and talking to relevant people. Consequently, it is important for Hallingdølen journalists to be able to search their archives that will provide a key source for information and backing data for new stories. Moreover, Hallingdølen journalists do not have specific tools (i) to search news from major Norwegian and international press as well as specific political web-sites, (ii) to find statistics, facts and to creatively explore numerical and quantified information about a news story in order to back it up with evidence in new and useful ways, and (iii) to discover and explore information different concept categories such as places, things, people and organisations relevant to the story at hand.

C. Requirements

Next, we outline key requirements that were elicited from the aforementioned scenarios.

1. Creatively explore different human angles in a news story based on the different people and roles associated with the story. This requirement focuses on presenting creative human angles to journalists, i.e. the names of identified individuals with distinct angles on the story, in a random order, to encourage the journalist to explore creatively the information about these angles.

2. Creatively explore numerical and quantified information about a news story in order to back it up with evidence in new and useful ways. This requirement highlights the need to visualise relevant quantifiable information.

3. Creatively explore the background events that underpin a story in order to discover a new angle to the story from its background. This requirement focuses on presenting background, causal events and information underpinning a story.

4. Rank articles by social media impact. Search social media for recent, relevant to a topic posts and tweets. This requirement refers to the capability to rank articles in provided sources by social media impact, where social media impact is related to the number of Likes / Shares.

5. Creatively explore unusual or comical information about a story as a means of using wit to report serious news. Suggesting funny angles to a story. This requirement explores both visual and textual quirky information related to a search topic.

III. TOOL DESIGN SPECIFICATION & IMPLEMENTATION

A. Architecture

For the construction of the architecture, we focused on key requirements and adopted the method of describing architectural components as outlined in [4]. We also put emphasis on providing rationales for all parts of the architecture and address cross-cutting concerns that will become important for subsequent development of all INJECT services and tools. The INJECT architecture is divided into three main layers, which are described next.

B. User interaction layer

To cater for the diverse needs of different types of journalism (such as freelancers or newsroom journalism), users can interact with the INJECT system through (i) Google Docs interface, which provides the INJECT functionalities through a add-on/sidebar and exploits Google Docs as the editor for creating, editing, storing and deleting news stories; (ii) CMS interface, which provides the INJECT functionalities through a TinyMCE plug-in/sidebar and exploits TinyMCE as the editor and associated popular CMSs such as Wordpress for creating, editing, storing and deleting news stories; (iii) Web-based interface, which provides the INJECT functionalities via a browser-displayed sidebar and allows journalists to use the INJECT functionalities and interact with desktop editing tools such as Microsoft Word using standard operating system capabilities.

C. Application layer

The application layer describes the key architecture components, their high-level functionality and interdependencies that will support journalists to undertake creative story-writing activities. Among the components of the Application Layer, Creative Search Service is the most innovative and is described in detail in the following section.

D. Data layer

The data layer provides access to internal and external resources and content repositories. Web and Social Media sources include over 300 public news sources and over 6,000,000 news articles that can be accessed programmatically for content search and retrieval by INJECT. Examples include the web sites of Le Monde, BBC as well as popular social media such as Twitter. News sources and archives, which are proprietary repositories of media and
articles used by the INJECT participating journalism SMEs and other third party repositories.

IV. CREATIVE SEARCH SERVICE

The creative search service was designed to manipulate input text automatically in order to generate search query terms that form the input to six creative strategies. Each of these six creative search strategies codified the practices that experienced journalists reported using to discover creative angles on news stories. To develop these strategies, semi-structured interviews with experienced journalists and digital experts in journalism were held to elicit candidate strategies. The candidate strategies were then validated with other experienced journalists, and extended and refined. Therefore, each of the resulting six strategies operationalized the elicited journalism expertise so that others could use it to discover possible creative angles on news stories. The six creative news angles are:

1. Backing and evidence: creatively explore numerical and quantified information about a news story in order to back it up with evidence in new and useful ways;
2. Individuals: creatively explore different human angles in a news story based on the different people and roles associated with the story;
3. Causal: creatively explore the background events that underpin a story in order to discover a new angle to the story from its background;
4. Quirky: creatively explore unusual or comical information about a story as a means of using wit to report serious news;
5. Ramifications: creatively explore information about the consequences of events and actions related to a news story in order to generate new and useful stories about future consequences;
6. Data visualizations: creatively explore different data sets and visualisations in order to generate new and useful ways about how to display information in a news story to encourage human insight.

Each of these 6 creative search strategies were based on the same manipulation of natural language text to make sense of then expand complex queries that implemented the different creative searches. All of the creative search services generate creative service queries in three steps. The first divides the inserted text into sentences that are then tokenized, part-of-speech tagged and modified to include each term’s morphological root (e.g. committed to commit) using the Brill Tagger [5]. The second applies increasingly sophisticated procedures to disambiguate each term by discovering its correct sense and tagging it with that sense using context knowledge from other terms in the query (e.g. defining a parliament to be a legislative assembly in certain countries rather than a card game) [18] [26]. The third implements
different creative search strategies that expand each term with other terms that have similar meanings to the tagged sense to search for and retrieve news information (e.g. the term parliament is synonymous with the terms congress, senate and house are then also included in the query based on the creative strategy).

Term senses are inferred automatically from WordNet, an online lexicon [25] that assigns senses to terms categorized as nouns, verbs, adjectives and adverbs. Each sense defines the meanings of a term, and WordNet organizes these senses into synonym sets that describe concepts with definitions or glosses, each of which contains a definition phrase composed of terms. The service uses this information about the implement 3 different creative search strategies:

1. **Synonym set expansion**: each disambiguated term is replaced with its synonym set, for example the verb commit is extended with the synonym set for the disambiguated sense #1 [perpetrate, pull] and noun government extended with the set for the disambiguated sense #1 [authorities, regime]. The strategy expands the query to retrieve news information about each object or action described in the inserted text;

2. **Hypernym expansion**: each disambiguated term is augmented by its direct hypernym. For example, the hypernym of the disambiguated term commit is act, so the revised query would include both terms. In our example, the service might retrieve news in-information about other forms of different government actions;

3. **Gloss words expansion**: each term is augmented with all terms in its gloss specified in WordNet. For example the sense #1 definition of the verb term commit is to perform an act, usually with a negative connotation. Hence the terms perform, act, negative and connotation are extracted and included in the extended query.

Each time that the service is invoked, it applies all three strategies to match sense-tagged query terms to a similar set of terms to return an unordered set of news articles or digital cartoons that achieved a threshold match score with the expanded search terms. Then, depending on the different creative search algorithm that is selected by the journalist, the service filters retrieved news articles and information using the terms specific to the strategy. For example, for the individual strategy, the service extracts names of each individual that has a Wikipedia entry from news articles. For the causal strategy, the service filters to retain matched articles with more than 500 words and a minimum threshold of keywords indicative of causal articles – terms such as cause, impact and studies – from sources such as the Economist and the New York Times. And for the backing and evidence strategy, the service filters to retain articles with a minimum threshold of quantity, measure and value keywords, for example Sterling, population and actual numbers. The full set of terms and other constraints that are used to filter news articles according to each constraint are not reported in this public deliverable.

Furthermore, the service has two different search modes – **Strict** and **Relaxed** – that the user can set using an interactive feature of the Sidebar during creative explorations. The default search mode is **Strict**, which generates queries composed of logical ANDs between all query terms shown at the top of the Sidebar. **Relaxed** mode generates queries composed of logical ORs between the terms at the top of the Sidebar and returns a broader set of generated angles.

V. TOOL WALKTHROUGH

To support primarily freelance journalists to undertake the pain-free exploration of information and ideas that is associated with creative thinking, INJECT’s digital creativity support was integrated as a fixed-width add-on sidebar on the right of the Google Docs editor window that the journalist could scroll to discover and examine creative angles generated by the tool beside the story that s/he was developing. Consider the following example: Imagine a fictional journalist who uses the INJECT Google Docs add-on to discover and examine creative angles for a story related to the on-going humanitarian crisis in Yemen. At any time s/he could highlight text written in the editor, then click the insert button on the right-hand top corner of the sidebar. In response, INJECT parses the highlighted text to extract stemmed nouns, verbs and proper names as candidate search terms, which the journalist could then edit before instructing INJECT to discover possible creative angles using search terms such as crisis, yemen and humanitarian, as shown in Figure 2. INJECT allows the journalist to review edit, add and delete each topic term directly in the top of sidebar.

Users can parameterise and control the following features and parameters:

- The publication date(s) of the news information that INJECT uses to generate creative guidance, to provide the journalist some control over generation and hence the creative guidance that the journalist receives;
- The different types of news source that INJECT uses to generate creative guidance – either from the 300 public news sources in the data layer or the news organisation’s local news archives, or both;
- The language of the news information that INJECT uses to generate creative guidance, to provide journalists with filters to control the language that the creative guidance is generated in. This language control feature emerged as an important requirement from journalists in different European countries;
The language of INJECT’s user interface, so that INJECT can offer journalists creativity support in the native languages of these journalists in order to reduce cognitive barriers to creative thinking with the news information and creative sparks.

The use of INJECT in evaluation mode. Evaluation mode was added to enable effective in-situ testing of the INJECT tool and services with consortium partners, to collect important testing data about the tool’s use.

To request creative guidance in INJECT, the journalist simply clicks one of the six creative search icons underneath the sidebar’s topic terms in INJECT’s strategy space. So for example, if the fictional journalist clicks the evidence and backing icon, INJECT uses the selected terms to search and retrieve news information sources that contain quantified information associated with a news story, and present relevant news stories in the sidebar. Figure 3 depicts an example fragment of the sidebar that presents news articles retrieved in

---

The Humanitarian Crisis in Yemen

The situation in Yemen is looking increasingly desperate with United Nations humanitarian agencies issuing an urgent new appeal for the Saudi-led coalition to fully lift its blockade on Yemen.

---

Figure 2. Google Docs plugin interface

The Humanitarian Crisis in Yemen

The situation in Yemen is looking increasingly desperate with United Nations humanitarian agencies issuing an urgent new appeal for the Saudi-led coalition to fully lift its blockade on Yemen.

---

Figure 3. The INJECT sidebar, showing news information retrieved using INJECT’s backing and evidence creative strategy, presented in interactive cards
response to the topic terms. If needed, s/he could request more people to search using the refresh button that is presented below the icons. The sidebar presents retrieved news articles in chunks of 12 with the ability to scroll down to access additional information.

In the sidebar, INJECT presents a simple explanation for the news information that it presents to the journalist. Furthermore, some of the elements of the explanation, such as publication dates and public/private news information sources, are interactive, and can be clicked on to change the elements. This interactivity was added to the explanation to encourage the journalist to play and explore with different news sources and information, as part of the creative discovery process. Furthermore, at any time, the journalist can request INJECT to provide a more in-depth explanation why it is presenting the news information that it is. INJECT generates adapted explanation texts, which it presents to the journalist in the sidebar space, as shown in the left pane of Figure 4.

In addition, if the journalist clicked quirky, INJECT searched for and retrieved political cartoons with matching names and captions, and presented each retrieved cartoon as a thumbnail image and caption in the sidebar and presented a creative spark, as shown in the right pane of Figure 4.

Figure 4. (Left) An example of INJECT’s more detailed explanation text, (right) an example of a cartoon’s thumbnail

VI. CONCLUSIONS

This paper described an innovative digital tool for facilitating creative exploration of information in journalism. The use of a 3-tier architecture has enabled cost-effective adaptation of the tool for use in different settings: the distinct interface layer has enabled new versions of INJECT to be delivered quickly for different text editor and web application environments; the separate application layer has been enabled these new versions of INJECT to have most if not all of the reasoning and interaction capabilities of INJECT at launch; and the separate data layer has enabled configurations of the INJECT tool to manipulate news and social media information from different private and public sources and written in different languages. Early evaluation of INJECT [14] by a limited number of journalists revealed that the tool was effective in supporting journalists discover new and useful angles on stories written with INJECT’s support, often quickly. However, the tool has been more effective when extended existing storylines rather than developed new ones. Journalist use of INJECT also provided more effective support for writing feature stories rather than news stories, although this finding might have been influenced by INJECT’s data layer, which was only composed of past news rather than current social media information. Results revealed that use of INJECT offered more support to journalists when discovering rather than examining. It appeared to contribute to journalists being more ambitious and to overcome biases by retrieving topic information from alternative news sources and discovering new stances from which to investigate news topics.

Our future work will refine INJECT’s capabilities by personalising its results according to the degree of the user’s familiarity with the tool. For example, it will offer greater transparency and shared control to less experienced users, while increasing the opaqueness for more experienced ones. Moreover, the development of anonymous user profiles by logging and processing user interactions with the tool will help tailor the tool’s results to the specific needs of the users. Finally, we will examine adding into the system generic profiles based on fictitious journalists personas, for example, according to the five categories of US journalists based on their social media habits and views of PR professionals: architects, promoters, hunters, observers and sceptics (https://www.prdaily.com/PRDailyEU/Articles/21493.aspx). These profiles will be selectable by the users and will allow them follow the information exploration tactics that different types of journalists would follow when tackling the story at hand.

ACKNOWLEDGMENT

This work is partly funded by the European Commission project H2020 INJECT (732278).

REFERENCES


