

Remixing Headlines for Context-Appropriate Flavor Text

Judith van Stegeren
Human Media Interaction
University of Twente
Enschede, The Netherlands
j.e.vanstegeren@utwente.nl

Mariët Theune
Human Media Interaction
University of Twente
Enschede, The Netherlands
m.theune@utwente.nl

Abstract—We describe a prototype of *Churnalist*, a headline generator for creating contextually-appropriate fictional headlines that can be used as flavor text in games. *Churnalist* creates new headlines by remixing existing headlines. It extracts seed words from free text input, searches for related words in a dataset of word embeddings and uses these words in the new headlines. The system requires no linguistic expertise or hand-coded language models from the user.

Index Terms—flavor text, headline generation, context, text generation, remixing, games

I. INTRODUCTION

Various video games, such as *Cities Skylines* [4], *Deus Ex: Human Revolution* [6] and *Cookie Clicker* [10] use fictional news to provide narrative context to the player. The fictional newspaper articles and headlines are an example of *flavor text*, i.e. text that is not essential to the main game narrative, but creates a feeling of immersion for the player. Flavor text is especially important for role-playing games and simulation games, as it gives the impression that the virtual world the player is interacting with is a living and breathing world. Writing flavor text is a time-consuming task for game writers. Text generation could be used to decrease the effort. We describe a prototype of *Churnalist*, a fictional headline generator that was created to support game writers in the task of writing flavor text.

II. HEADLINE GENERATION

Headline generation is often seen as a document summarization task, where headline generators take a full article text as input and return a headline that describes the most salient theme or the main event of the text. Existing approaches for headline generation are based on rules [5], statistics [1] and machine learning [3], [9], with the latter winning in popularity in recent years.

An approach that is similar to ours is that of *HeadyLines* [7], a headline generator that focuses on the creative side of writing headlines. It can be used to support editors in their task of writing catchy news headlines. Given a newspaper article text as input, it extracts the most important words from the text

This research is part of DATA2GAME (project number 055.16.114), funded by The Netherlands Organisation for Scientific Research (NWO).

and uses these as seed words for generating a set of variations on well-known lines, such as movie names and song lyrics.

One difference between *Churnalist* and typical headline generators is that instead of summarizing the input text, our system creates headlines that fit the context of the input, i.e. new headlines are related to the input text, but do not describe it.

III. CONTEXT

The objective of our headline generation system is to take existing headlines and adapt them to a new context: the context of the input text. Generating text that appears to be coherent and context-appropriate has been studied before, such as in computational creativity [11] and procedural generation for games [8]. Notably, Veale [11] observed that users will generously attribute meaning to a computer generated text when it has a familiar form. For this he coined the phrase *charity of interpretation*. *Churnalist* reuses words from the input text in the generated headlines. By reusing these words in a form that is familiar to the reader, i.e. headlines, we want to exploit the charity of interpretation effect in our readers to evoke the context of the input text.

IV. SYSTEM DESCRIPTION

Churnalist is meant for creating headlines as flavor text for video games. The input can be anything, from a snippet of game narrative to lines of character dialogue. As long as the input text contains noun phrases that are representative for the narrative context, *Churnalist* can use the noun phrases to create headlines. The advantage of this approach is that *Churnalist* can be used for different games and topics.

Headline generation in *Churnalist* takes place in three steps. In the first step, the system reads the input text and extracts the most important words (noun phrases), which will be used as *seed words*. In the second step, *Churnalist* expands the list of seed words with a set of loosely related words by searching for related words in *FastText* [2], a pre-trained vector space of word embeddings. The system calculates a vector for the head noun of each noun phrase and tries to find its closest neighbours in the vector space. These neighbouring words are added to the list of seed words. The final step uses word substitution to create a new headline. A random headline from

“You are the system administrator of SuperSecure Ltd, a hosting company. At four o’clock in the afternoon, your manager storms in. Apparently, there has been a break-in in your computer network. The CEO has been receiving anonymous emails from a hacker that demands a payment of \$100,000 before midnight.”

Fig. 1. Example input text. The system will use noun phrases (underlined) as seed words.

Seed word	suggestions
company	subsidiary, webcompany
CEO	executive, shareholder, entrepreneur, investor
manager	teammanger
hackers	hacktivists, cybercriminals, scammers

Fig. 2. Seed words and examples of suggestions for related words.

a database is parsed by SpaCy’s dependency parser¹, after which the system substitutes the subject of the headline with a seed word. At the end of every step in the generation process, the user can filter the output, thus fine-tuning the nouns and noun phrases that are used in later generation steps.

Even though the separate steps in the system are relatively simple, implementing a generator that can fulfil the task robustly shows some of the practical challenges of text generation. For example, the system should be able to parse the special grammar of headlines, handle out-of-vocabulary words in the input and insert seed words in the headline with the right inflection.

V. EXAMPLE

Figure 1 shows a short narrative text about computer security, as an example of representative input text for Churnalist. The system begins with extracting all noun phrases. The head nouns of all noun phrases (e.g. ‘network’ in the noun phrase ‘computer network’) are selected as seed words. Churnalist then searches for words related to these seed words. After filtering out non-existing words from the results (necessary because FastText contains a lot of typographical errors), we get the related words shown in Figure 2.

In the final generation step, the system picks a random headline from a headline database, which will be used as the starting point for the output. If the random headline is ‘Police urgently try to find parents of toddler found in street’, the dependency parser will flag ‘police’ as the subject of this sentence. Churnalist randomly selects ‘hacker’ as the seed word for the new headline and inflects it to the plural form, ‘hackers’, since ‘police’ is plural in the original headline. The headline’s subject is substituted with the inflected seed word, creating the new headline ‘Hackers urgently try to find parents of toddler found in street’. See Figure 3 for more examples of the headlines that Churnalist generates.

¹SpaCy 2.0.16, <https://www.spacy.io>

“Revealed: system administrator is a convicted people smuggler”
 “Hosting company has edge over Trump on budget negotiations, CBS News poll shows”
 “Computer network issues ice warning as snow hits UK”
 “Hacker loses latest legal bid over driver rights”
 “Hosting company revises cause of escape room fire that killed 5 girls”
 “CEO’s threat to block government’s tax without second Brexit referendum”

Fig. 3. Generated headlines for the input text in Figure 1. Seed words are underlined.

VI. FUTURE WORK

In the headlines that Churnalist currently generates, there is often a mismatch between the selected seed word and the verb of the sentence, e.g. ‘payments write emotional plea’. Instead of a random headline, Churnalist should pick a headline for substitution that already has a link with the chosen seed word. We want to use semantic networks, such as ConceptNet and WordNet, to supplement the word embeddings so that Churnalist suggests only valid words to the user when expanding the list of seed words. Additionally, Churnalist could use semantic networks to only suggest words that fit the headline context in terms of gender, number, and animacy. After further improving the system, we want to compare Churnalist’s performance with a state-of-the-art neural headline generator, using the same game texts as input.

REFERENCES

- [1] Michele Banko, Vibhu O Mittal, and Michael J Witbrock. Headline generation based on statistical translation. In *Proceedings of the 38th Annual Meeting of ACL*, pages 318–325. Association for Computational Linguistics, 2000.
- [2] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association of Computational Linguistics*, 5:135–146, 2017.
- [3] Carlos A Colmenares, Marina Litvak, Amin Mantrach, and Fabrizio Silvestri. Heads: Headline generation as sequence prediction using an abstract feature-rich space. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 133–142, 2015.
- [4] Colossal Order. *Cities: Skylines*. Game [PC], 2017. Paradox Interactive, Stockholm, Sweden.
- [5] Bonnie Dorr, David Zajic, and Richard Schwartz. Hedge Trimmer: A parse-and-trim approach to headline generation. In *Proceedings of the HLT-NAACL 03 Text Summarization Workshop*, pages 1–8, 2003.
- [6] Eidos Montral. *Deus Ex: Human Revolution*. Game [PC], 2011. Square Enix, Shinjuku, Tokyo, Japan.
- [7] Lorenzo Gatti, Gözde Özbal, Marco Guerini, Oliviero Stock, and Carlo Strapparava. Heady-lines: A creative generator of newspaper headlines. In *Companion Publication of the 21st International Conference on Intelligent User Interfaces, IUI 2016*, pages 79–83, 2016.
- [8] Jason Grinblat and C. Brian Bucklew. Subverting historical cause & effect: generation of mythic biographies in Caves of Qud. In *Proceedings of the 12th International Conference on the Foundations of Digital Games*, pages 1–7, New York, NY, USA, 2017. ACM, ACM.
- [9] Shi-Qi Shen, Yan-Kai Lin, Cun-Chao Tu, Yu Zhao, Zhi-Yuan Liu, Mao-Song Sun, et al. Recent advances on neural headline generation. *Journal of Computer Science and Technology*, 32(4):768–784, 2017.
- [10] Julien “Orteil” Thiennot. *Cookie Clicker*. Game [PC/Browser], August 2013. <http://orteil.dashnet.org/cookieclicker/>. Played September 2018.
- [11] Tony Veale. The shape of tweets to come: Automating language play in social networks. *Multiple Perspectives on Language Play*, 1:73–92, 2016.