



Generating headlines as flavour text for a dilemma-based game

CLIN29, Groningen

Judith van Stegeren

January 31, 2019

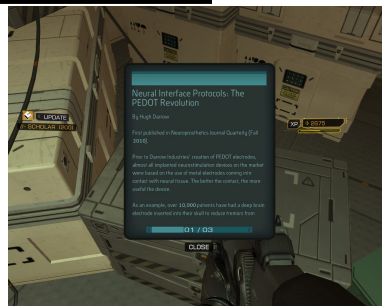
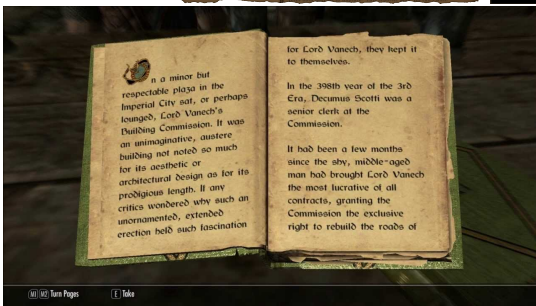


Flavour text



Trading cards, Magic the Gathering (Wizards of the Coast)

Flavour text



Flavour text in RPG-games *Bioshock*, *Diablo III*, *Skyrim* and *Deus Ex: Human Revolution*

Flavour text

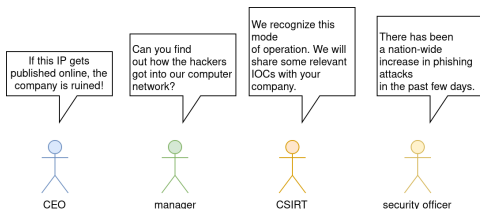
Flavour text is text that provides extra information about characters, objects, settings etc.

While appropriate to the game's story concept, it usually has no effect on the mechanics of the game, but instead serves to add depth or background, providing a form of context.

DATA2GAME

- ▶ Player modeling, sensor technology and natural language generation for making adaptive serious games
- ▶ <https://www.data2game.nl>

Dilemma game



You are the system administrator of SuperSecure Ltd, a hosting company.

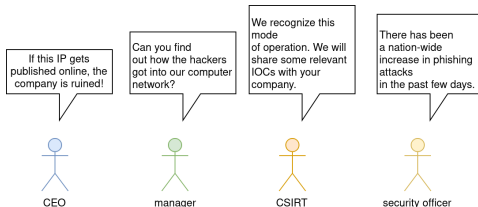
At four o'clock in the afternoon, your manager storms in. Apparently, there has been a break-in in your computer network. The CEO has been receiving anonymous emails from a hacker that demands a payment of \$100,000 before midnight. If SuperSecure does not pay, they threaten to publish sensitive company documents online. The manager is worried, since the hacker claims to possess important intellectual property of one of SuperSecure's clients.

However, you cannot find any traces of malware within the network and you find nothing suspicious in your log files.

Hire an expensive external security consultant to check your network logs

Reassure the manager and the CEO: the hacker is probably bluffing

Dilemma game



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. If SuperSecure does not pay, they threaten to publish sensitive company documents online. The manager is worried, since the hacker claims to possess important intellectual property of one of SuperSecure's clients.

However, you cannot find any traces of malware within the network and you find nothing suspicious in your log files.

Hire an expensive external security consultant to check your network logs

Reassure the manager and the CEO: the hacker is probably bluffing

News

Minister warns for phishing attacks

How hackers threaten Arctic Ocean life

11 facts which will make you a computer network expert

Trump delays Congress speech over active malware

The trick that could help you avoid malicious activity

Maduro cuts ties after US backs hackers

Goal for the system

Game text as input → set of generated headlines as output

Input: game text

“You are the system administrator of SuperSecure Ltd, a hosting company. At four o'clock in the afternoon, your manager storms in. Manager: Can you find out how the hackers got into our computer network? CSIRT: We recognize this mode of operation. We will share some relevant IOCs with your company. Can you contact us if you have finished your forensical analysis? Security officer: There has been a nation-wide increase in phishing attacks in the past few days. System administrator: I can't find any traces of active malware on our Windows server. I will check the network log files for malicious activity.”

Output: headlines

- ▶ Minister warns for phishing attacks
- ▶ How hackers threaten Artic Ocean life
- ▶ 11 facts which will make you a computer network expert
- ▶ Trump delays Congress speech over active malware
- ▶ The trick that could help you avoid malicious activity
- ▶ Maduro cuts ties after US backs hackers

Goal for the system

Game text as input → set of generated headlines as output

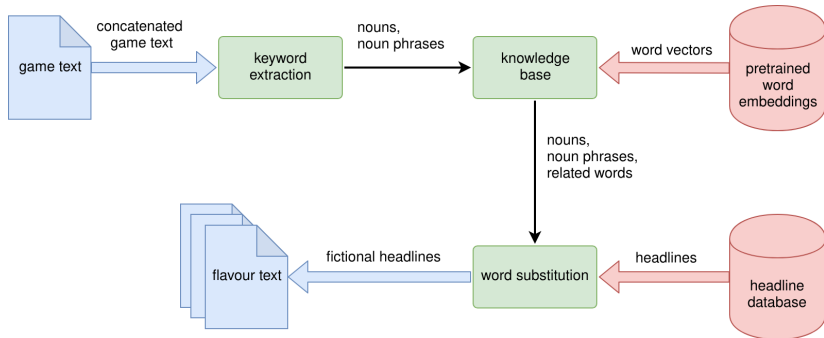
Input: game text

“You are the system administrator of SuperSecure Ltd, a hosting company. At four o'clock in the afternoon, your manager storms in. Manager: Can you find out how the **hackers** got into our **computer network**? CSIRT: We recognize this mode of operation. We will share some relevant IOCs with your company. Can you contact us if you have finished your forensical analysis? Security officer: There has been a nation-wide increase in **phishing attacks** in the past few days. System administrator: I can't find any traces of **active malware** on our Windows server. I will check the network log files for **malicious activity**.”

Output: headlines

- ▶ Minister warns for **phishing attacks**
- ▶ How **hackers** threaten Artic Ocean life
- ▶ 11 facts which will make you a **computer network** expert
- ▶ Trump delays Congress speech over **active malware**
- ▶ The trick that could help you avoid **malicious activity**
- ▶ Maduro cuts ties after US backs **hackers**

Churnalist system overview



Keyword extraction

Input: game text

"You are the system administrator of SuperSecure Ltd, a hosting company. At four o'clock in the afternoon, your manager storms in. Manager: Can you find out how the hackers got into our computer network? CSIRT: We recognize this mode of operation. We will share some relevant IOCs with your company. Can you contact us if you have finished your forensic analysis? Security officer: There has been a nation-wide increase in phishing attacks in the past few days. System administrator: I can't find any traces of active malware on our Windows server. I will check the network log files for malicious activity."

Output: list of nouns and noun phrases

- ▶ administrator → system administrator
- ▶ company → hosting company
- ▶ manager
- ▶ hackers
- ▶ network → computer network
- ▶ operation
- ▶ analysis → forensic analysis
- ▶ (...)

Knowledge base

To increase the size of our word list, we query a pre-trained word embeddings model¹ for words close to nouns in the word list.

Input: list of nouns

- ▶ administrator
- ▶ company
- ▶ manager
- ▶ hackers
- ▶ network
- ▶ operation
- ▶ analysis
- ▶ phishing

¹E. Grave, P. Bojanowski, P. Gupta, A. Joulin, T. Mikolov, *Enriching Word Vectors with Subword Information*

Knowledge base

To increase the size of our word list, we query a pre-trained word embeddings model¹ for words close to nouns in the word list.

Input: list of nouns

- ▶ administrator
- ▶ company
- ▶ manager
- ▶ hackers
- ▶ network
- ▶ operation
- ▶ analysis
- ▶ phishing

Output: for every noun, a list of related nouns

- ▶ subsidiary
- ▶ subsidiaries
- ▶ ecompanies
- ▶ corporation
- ▶ firm
- ▶ subcompanies
- ▶ ltd*
- ▶ stockholder
- ▶ manufacturer

¹E. Grave, P. Bojanowski, P. Gupta, A. Joulin, T. Mikolov, *Enriching Word Vectors with Subword Information*

Substitution

Original headline

Mnangagwa promises investigation of brutal Zimbabwe crackdown

Democrats boycott meeting with Trump at White House to discuss government shutdown

Airport drone disruption: All major UK airports to have military-grade protection

Tottenham reveal SHOCKING news about transfer strategy

New headline

System administrator promises investigation of brutal SuperSecure Ltd crackdown

Democrats boycott meeting with Trump at White House to discuss hackers

Hacker drone disruption: all major UK airports to have military-grade protection

Manager reveals SHOCKING news about phishing emails

Future work: system

Most challenges are related to *engineering*, not linguistics.

1. keyword extraction: extract proper names
2. keyword extraction: use tf-idf for filtering nouns
3. keyword extraction: ask human to filter nouns and names
4. knowledge base: use Wordnet and FastText data in tandem
5. substitution step: filter headlines from DB
6. substitution step: implement generate-and-test strategy (n-grams)
7. substitution step: substitute subject and object of headline
8. substitution step: post-processor for cosmetic improvements

Future work: evaluation

1. compare output to baseline and output of neural headline generation system
2. evaluate output text with humans for 'relevance' property
3. evaluate system with game developers for evaluating usefulness

Questions?



- ▶ Judith van Stegeren
- ▶ j.e.vanstegeren@utwente.nl
- ▶ <https://twitter.com/jd7h>
- ▶ <https://www.judithvanstegeren.com>
- ▶ <https://www.data2game.nl>

