

Clickbait Spoiling via Question Answering and Passage Retrieval

22nd–27th May, 2022



Matthias Hagen¹



Maik Fröbe¹



Artur Jurk¹



Martin Potthast²

¹Martin-Luther-Universität Halle-Wittenberg
first.last@informatik.uni-halle.de

²Leipzig University
martin.potthast@uni-leipzig.de

Clickbait Spoiling via Question Answering and Passage Retrieval

Spoilers: The Dark Side



Clickbait Spoiling via Question Answering and Passage Retrieval

Spoilers: The Dark Side



- ❑ A spoiler is a short summary that reveals details of a plot
- ❑ Accidentally seeing a spoiler may ruin the experience
- ❑ Wikipedia had spoiler alerts until 2007

Clickbait Spoiling via Question Answering and Passage Retrieval

Spoilers: The Dark Side



- ❑ A spoiler is a short summary that reveals details of a plot
- ❑ Accidentally seeing a spoiler may ruin the experience
- ❑ Wikipedia had spoiler alerts until 2007

Clickbait Spoiling via Question Answering and Passage Retrieval

Spoilers: The Light Side



Lifehacker ✓
@lifehacker



How to keep your workout clothes from stinking:
lifelifehack.com/57Y0uEZ



Clickbait [Potthast et al., 2018]

- ❑ Social-media posts that create a curiosity gap to generate clicks
- ❑ Resolution is usually ordinary or trivial

Clickbait Spoiling via Question Answering and Passage Retrieval

Spoilers: The Light Side



Lifehacker ✓
@lifehacker



How to keep your workout clothes from stinking:
lifelifehack.com/57Y0uEZ



Saved You A Click
@SavedYouAClick



Wash them.

Clickbait [Potthast et al., 2018]

- ❑ Social-media posts that create a curiosity gap to generate clicks
- ❑ Resolution is usually ordinary or trivial

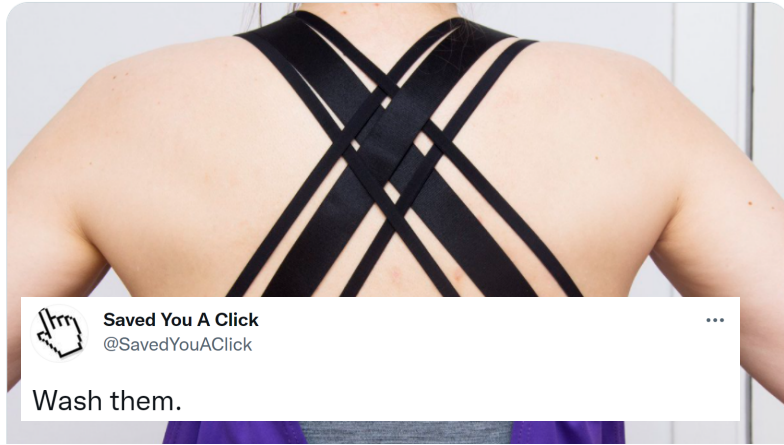
Clickbait Spoiling via Question Answering and Passage Retrieval

Spoilers: The Light Side



Lifehacker
@lifehacker

How to keep your workout clothes from stinking:
lifelifehack.com/57Y0uEZ



Gizmodo
@Gizmodo

What the 'someone is typing' bubbles in messaging apps actually mean gizmodo.com/jodfFXV



Clickbait [Potthast et al., 2018]

- ❑ Social-media posts that create a curiosity gap to generate clicks
- ❑ Resolution is usually ordinary or trivial

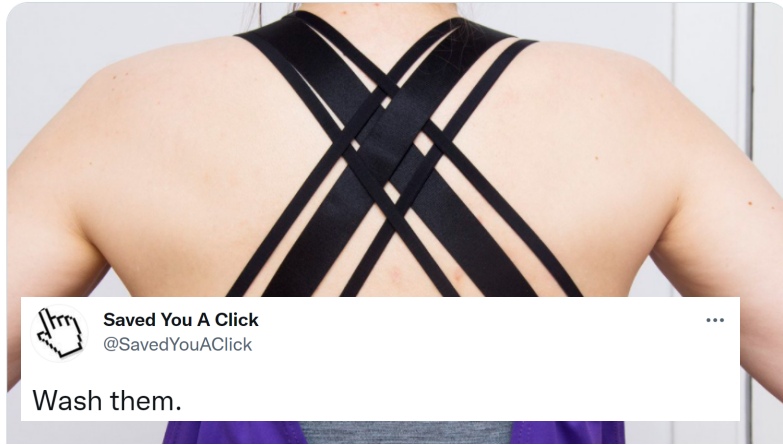
Clickbait Spoiling via Question Answering and Passage Retrieval

Spoilers: The Light Side



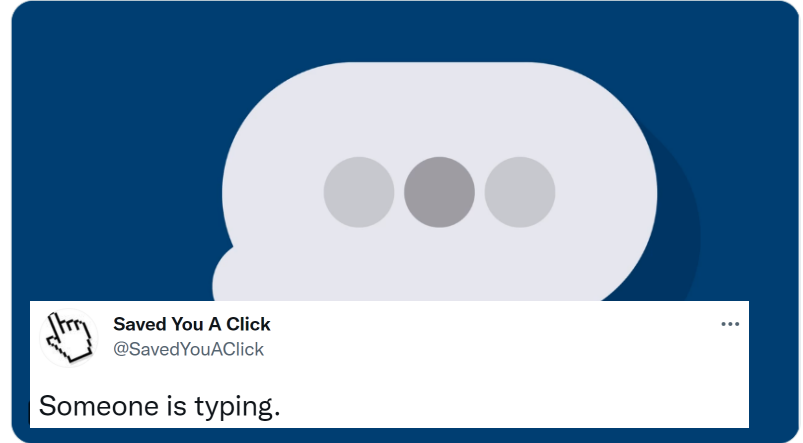
Lifehacker
@lifehacker

How to keep your workout clothes from stinking:
lifelifehack.com/57Y0uEZ



Gizmodo
@Gizmodo

What the 'someone is typing' bubbles in messaging apps actually mean gizmodo.com/jodfFXV



Clickbait [Potthast et al., 2016]

- ❑ Social-media posts that create a curiosity gap to generate clicks
- ❑ Resolution is usually ordinary or trivial

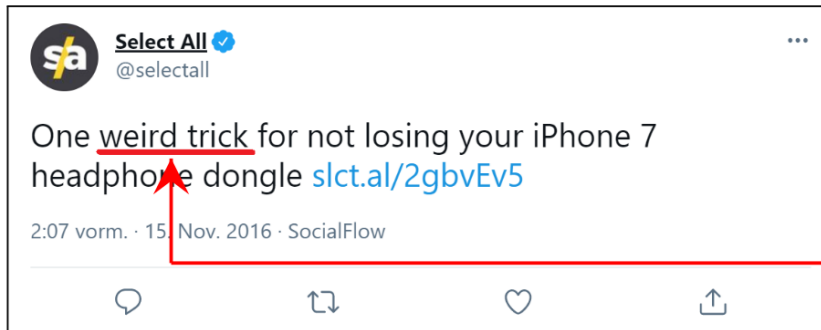
Clickbait Spoiling


- ❑ Generating a short text that satisfies the curiosity gap

Webis Clickbait Spoiling Corpus 2022

Corpus Construction

Goal: 5,000 spoilable clickbait posts with high quality annotations



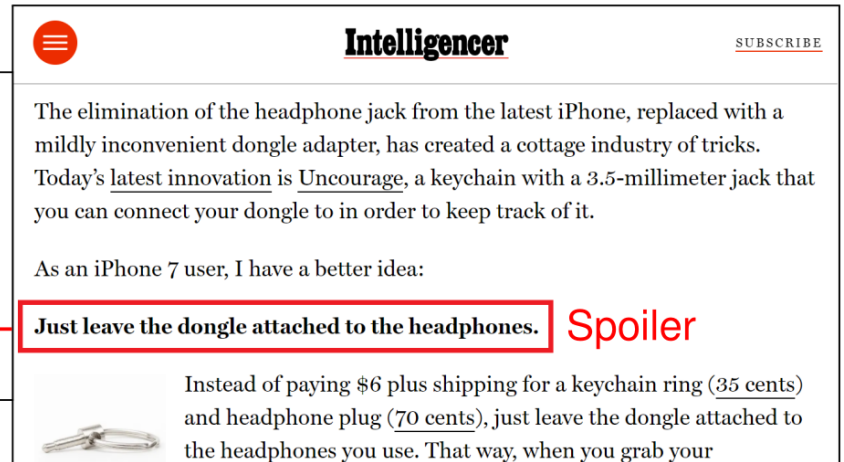
 **Select All** ✓
@selectall


One weird trick for not losing your iPhone 7 headphone dongle slct.al/2gbvEv5

2:07 vorm. · 15 Nov. 2016 · SocialFlow

⋮

🗨️ ↺️ ❤️ ↗️




 **Intelligencer** SUBSCRIBE

The elimination of the headphone jack from the latest iPhone, replaced with a mildly inconvenient dongle adapter, has created a cottage industry of tricks. Today's latest innovation is Uncourage, a keychain with a 3.5-millimeter jack that you can connect your dongle to in order to keep track of it.

As an iPhone 7 user, I have a better idea:

Just leave the dongle attached to the headphones. **Spoiler**

 Instead of paying \$6 plus shipping for a keychain ring (35 cents) and headphone plug (70 cents), just leave the dongle attached to the headphones you use. That way, when you grab your

Webis Clickbait Spoiling Corpus 2022

Corpus Construction

Goal: 5,000 spoilable clickbait posts with high quality annotations

The image shows a tweet from the account 'Select All' (@selectall) dated 2:07 vorm. · 15 Nov. 2016. The tweet text is 'One weird trick for not losing your iPhone 7 headphone dongle slct.al/2gbvEv5'. A red arrow points from the word 'headphone' in the tweet to a red box in the article content. The article is from 'Intelligencer' and discusses a 'dongle adapter' for iPhone 7. The spoiler text, 'Just leave the dongle attached to the headphones.', is highlighted in a red box and labeled 'Spoiler' in red text. Below the spoiler is a small image of a metal keychain ring and a text description of the 'better idea'.

Starting point:

- ❑ 5 social media accounts that manually spoil clickbait
- ❑ Webis Clickbait-17 dataset

560 hours annotation effort:

- ❑ Manual post selection, main content extraction, and spoiler identification
- ❑ 3 types of spoilers: Phrase, Passage, Multi

Webis Clickbait Spoiling Corpus 2022

Spoiler Types

❑ Phrase

- ≤ 5 consecutive words
- Often entities and nouns

Someone is typing.

❑ Passage

- > 5 consecutive words
- Often Descriptions

**Just leave the
dongle attached to
the headphones.**

❑ Multipart

- Lists closing multiple curiosity gaps

**1. Set Your Alarm with
Precision
2. Write Down 1 Daily
Intention**

Webis Clickbait Spoiling Corpus 2022

Spoiler Types

❑ Phrase

- ≤ 5 consecutive words
- Often entities and nouns

Someone is typing.

❑ Passage

- > 5 consecutive words
- Often Descriptions

**Just leave the
dongle attached to
the headphones.**

❑ Multipart (future work)

- Lists closing multiple curiosity gaps

**1. Set Your Alarm with
*Precision***
**2. Write Down 1 Daily
Intention**

Webis Clickbait Spoiling Corpus 2022

Corpus Overview

Spoiler	Entries	Corpus splits		
		Train	Val.	Test
Phrase	2,125	1,367	335	423
Passage	1,999	1,274	322	403
Multipart	876	559	143	174
Sum	5,000	3,200	800	1,000

Webis Clickbait Spoiling Corpus 2022

Corpus Overview

Spoiler	Entries	Corpus splits		
		Train	Val.	Test
Phrase	2,125	1,367	335	423
Passage	1,999	1,274	322	403
Multipart	876	559	143	174
Sum	5,000	3,200	800	1,000

Supported Subtasks

Spoiler type classification

- ❑ Input: Post, Article
- ❑ Output: Phrase, Passage, Multipart

Spoiler generation

- ❑ Input: Post, Article, Spoiler type
- ❑ Output: text that satisfies the curiosity gap of the post

Spoiler Type Classification

Feature-Based Models

- ❑ Models: Naïve Bayes, Logistic Regression, SVM
- ❑ Features: tf / $tf \cdot idf$ of uni/bi-grams from the post and the linked web page
- ❑ Chi-square feature selection

Transformer Models

- ❑ Models: BERT, DeBERTa, and RoBERTa
- ❑ Input: Post + main content of the linked page

Spoiler Type Classification

Feature-Based Models

- ❑ Models: Naïve Bayes, Logistic Regression, SVM
- ❑ Features: tf / $tf \cdot idf$ of uni/bi-grams from the post and the linked web page
- ❑ Chi-square feature selection

Transformer Models

- ❑ Models: BERT, DeBERTa, and RoBERTa
- ❑ Input: Post + main content of the linked page

Results (Phrase vs. Passage)

Model	Accuracy
Naïve Bayes	67.07
SVM	69.61
Log. Regression	70.10
BERT	76.27
DeBERTa	79.06
RoBERTa	80.39

Spoiler Generation

Effectiveness (1)

- Dedicated models per spoiler type
 - Question answering for phrase spoilers
 - Passage retrieval for passage spoilers
- Pilot study with 20 models
- Question answering models first fine-tuned on SQuAD then on our corpus

Type	Model	Phrase Spoilers		Passage Spoilers	
		BERTScore	P@1	BERTScore	P@1
Question Answering	BERT (baseline)	71.06	66.67	44.38	42.43
	DeBERTa-large	77.03	75.65	51.06	54.84
	RoBERTa-large	74.81	72.58	49.99	53.85
Passage Retrieval	BM25 (baseline)	19.94	8.27	34.71	4.22
	monoBERT	20.66	42.08	36.58	26.05
	monoT5	20.98	43.97	36.70	29.03

Spoiler Generation

Effectiveness (2)

- Ablation study: Do we need spoiler type classification?

Two-Step Pipeline		End-to-End Effectiveness	
Type Classification	Spoiler Generation	BERTScore	P@1
RoBERTa	DeBERTa	59.18	63.44
None	DeBERTa	60.43	64.16
Oracle	DeBERTa	64.36	65.50

Conclusion

New Task: Clickbait spoiling

- ❑ Present a short text that satisfies the curiosity induced by a clickbait post

Webis Clickbait Spoiling Corpus 2022

- ❑ First large resource for clickbait spoiling

First baselines for clickbait spoiling

Code and data: github.com/webis-de/acl-22

Conclusion

New Task: Clickbait spoiling

- ❑ Present a short text that satisfies the curiosity induced by a clickbait post

Webis Clickbait Spoiling Corpus 2022

- ❑ First large resource for clickbait spoiling

First baselines for clickbait spoiling

Code and data: github.com/webis-de/acl-22

Future work

Explore other approaches to clickbait spoiling

- ❑ Paraphrase a clickbait post so that it contains its own spoiler

Multipart spoilers

- ❑ Summarization models to select the different parts of the document

Conclusion

New Task: Clickbait spoiling

- ❑ Present a short text that satisfies the curiosity induced by a clickbait post

Webis Clickbait Spoiling Corpus 2022

- ❑ First large resource for clickbait spoiling

First baselines for clickbait spoiling

Code and data: github.com/webis-de/acl-22

Future work

Explore other approaches to clickbait spoiling

- ❑ Paraphrase a clickbait post so that it contains its own spoiler

Multipart spoilers

- ❑ Summarization models to select the different parts of the document

Thank You!