

BaitBuster: Destined to Save You Some Clicks

Md Main Uddin Rony¹, Naeemul Hassan¹, Mohammad Yousuf²
¹The University of Mississippi, ²The University of Oklahoma



Motivation



Clickbaiting has become widespread

- All types of media practice clickbait.
- It has become a source of easy revenue.

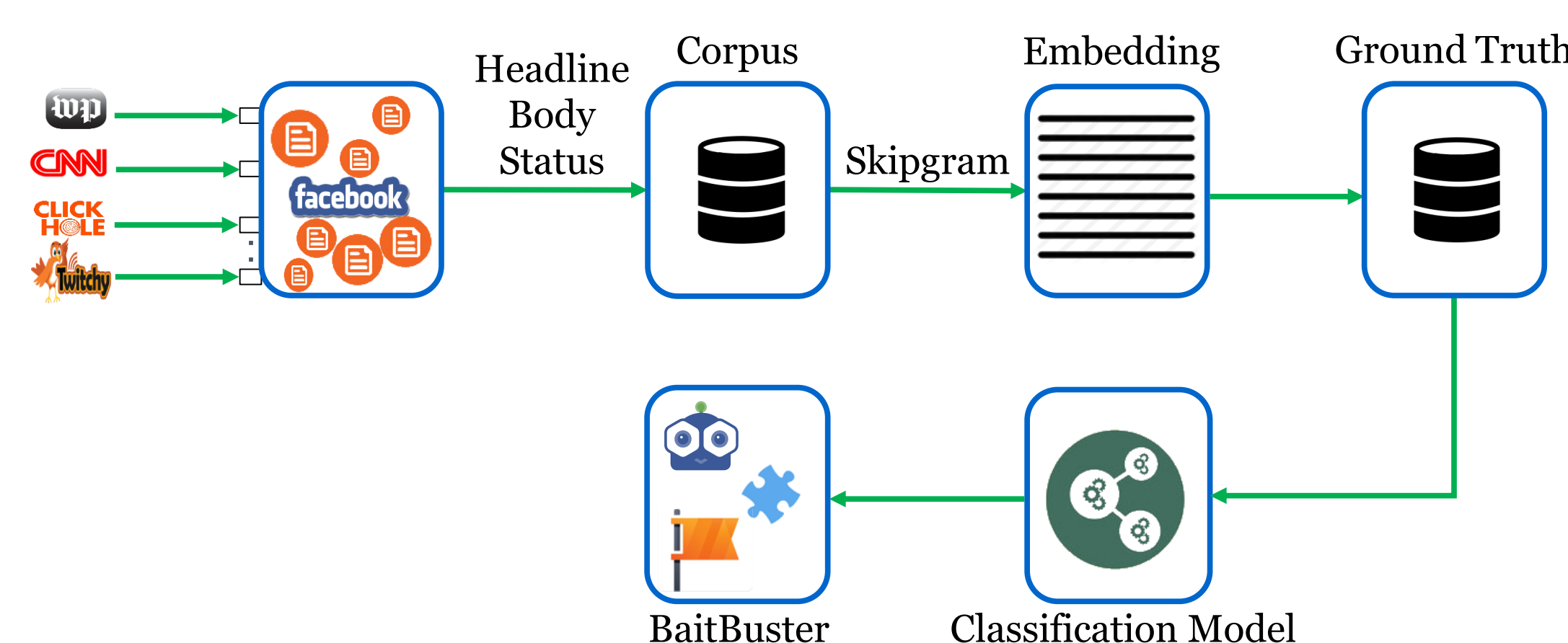
Negative impact on media ecosystem

- Damages media-user relationship.
- Depletes media brand value.

Lack of comprehensive solution

- A little but inefficient approaches to fight against this ill practice.
- Solutions depending on checking the headline only are not sufficient.

Problem Formulation



Two categories of headlines

Clickbait

- Eek! What's Lurking in the Shadows?! I Have to Know!*
- I Left My Daughter And THIS Happened!*

Non-Clickbait

- America's democracy has become illiberal*
- Twitter Expands Fight Against Abuse*

Goal

- Developing a clickbait detection model.
- Building a clickbait solution framework to improve the web surfing experience.

Data Collection

Ground Truth

- 32,000 manually labeled news headlines. [1]

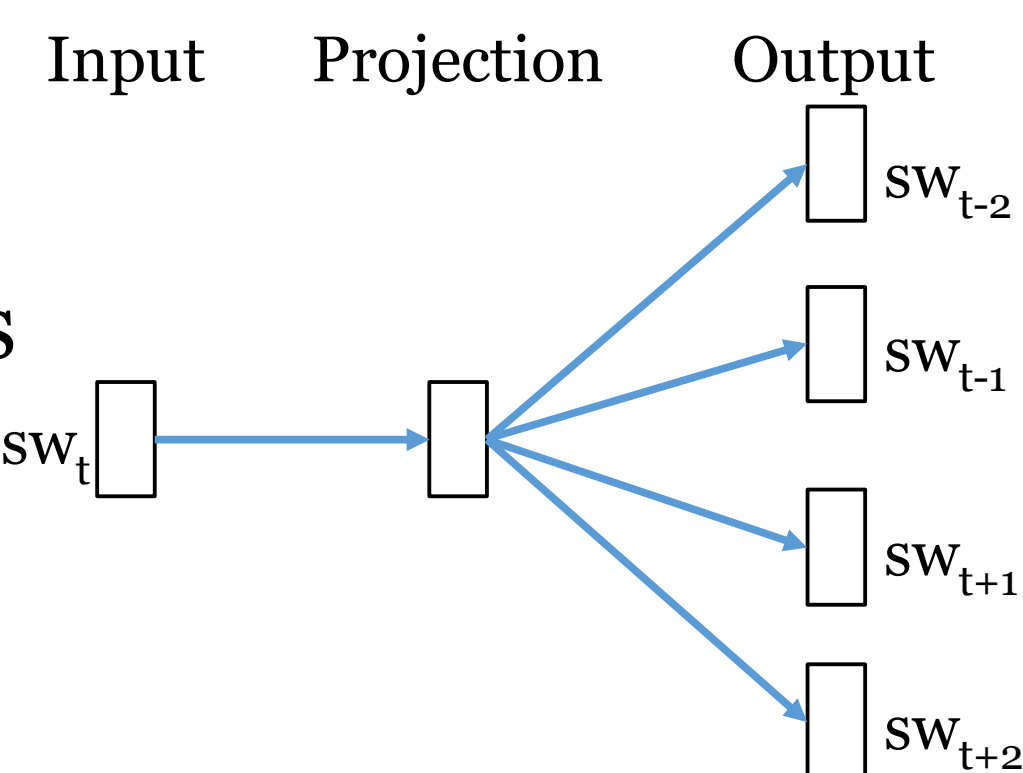
Media Corpus

- About 1.7 million Facebook posts.
- From 68 mainstream and 85 unreliable media.
- Considered headlines, bodies, and the statuses.

Media	Category	Link	Video	Total
Mainstream	Broadcast	324028	32924	356952
	Print	516713	14129	530842
Unreliable	Clickbait	371834	4099	375933
	Conspiracy	309122	5841	314963
	Junk Science	51923	649	52572
	Satire	41046	151	41197
Total		1614666	57793	1672459

Clickbait Detection

- Distributed subword embedding was used to create sentence vectors.
- Pre-trained embeddings using 477236 words of the Media Corpus.
- Used Softmax function as the classifier.



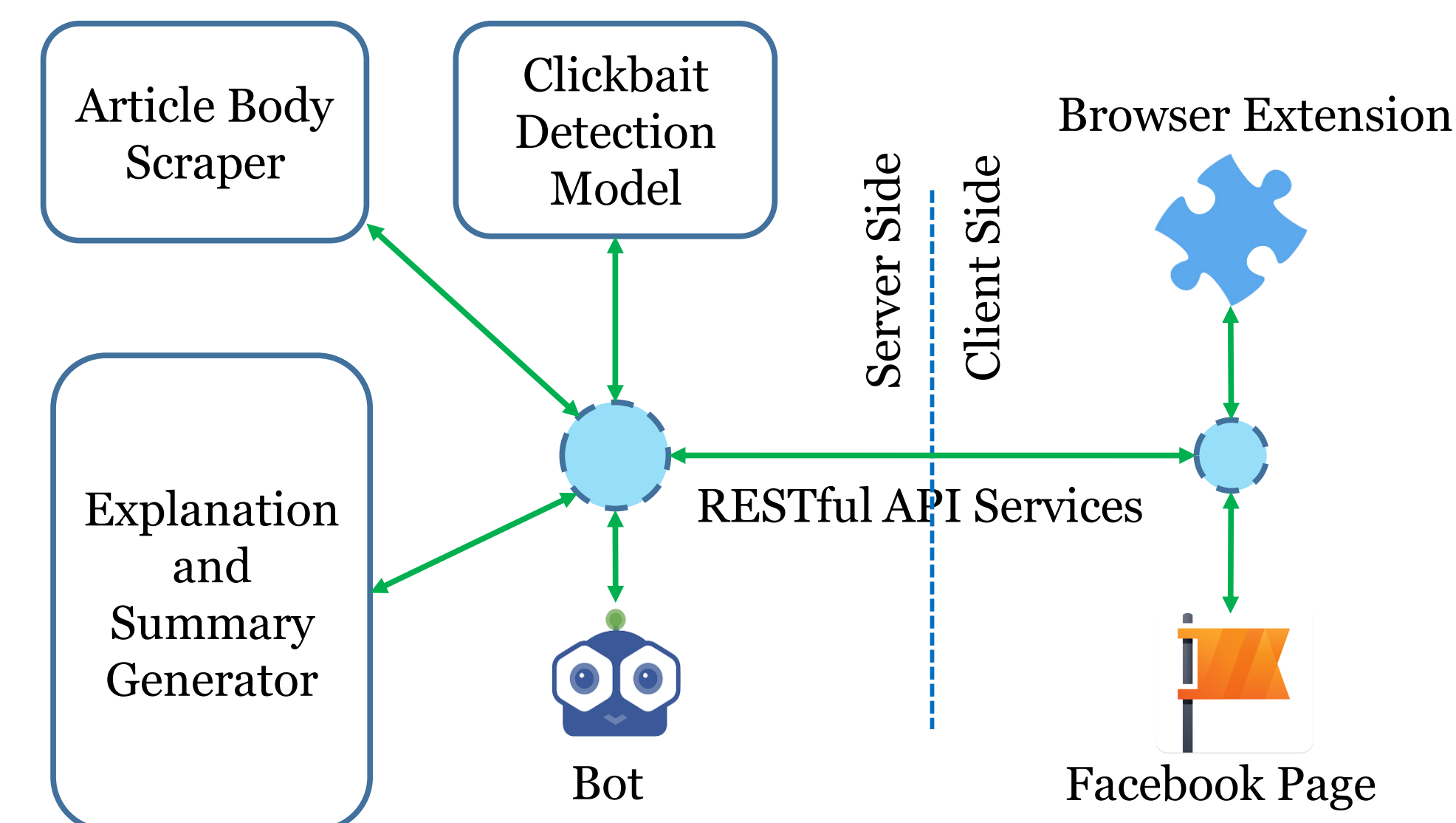
	Method	Precision	Recall	F-score	Accuracy
Without Pre-trained Vectors	Chakraborty et al. [1]	0.95	0.90	0.93	0.93
	Ground Truth	0.976	0.975	0.975	0.976
With Pre-trained Vectors	Anand et al. [2]	0.984	0.978	0.982	0.982
	Ground Truth + Google word2vec	0.977	0.977	0.977	0.976
	Ground Truth + Media Corpus	0.983	0.983	0.983	0.983

BaitBuster

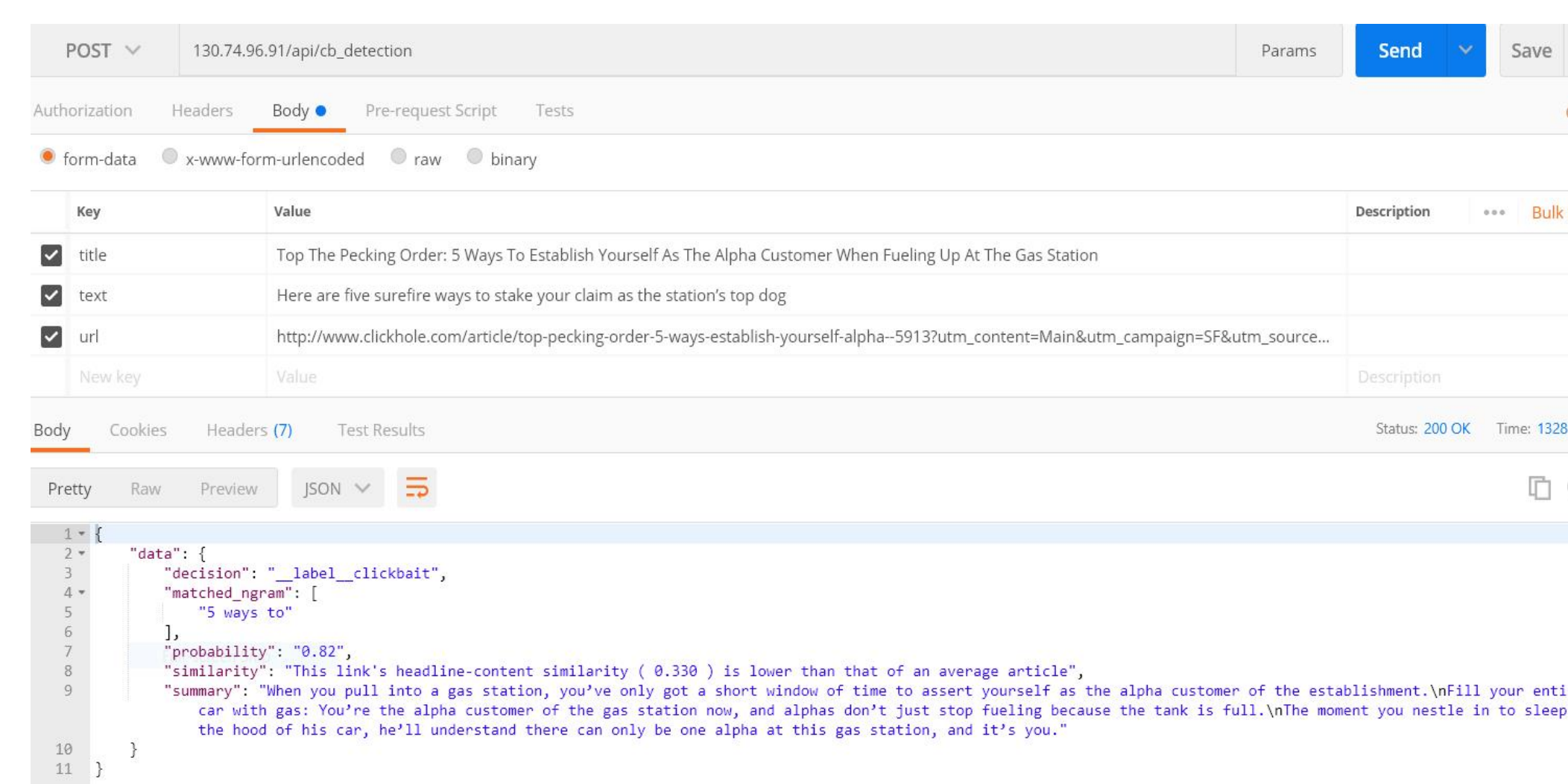
Overview

- Powered by clickbait detection model with the accuracy of 98.3%.
- Provides brief explanation behind clickbait decision.
- Allows users to read the brief summary of the content without leaving the current page.
- Collects user feedback on clickbait decision.
- Social bot regularly publishes automatically generated report about contemporary clickbait contents.

System Architecture



Clickbait Detection API



Clickbait & Media Reliability

Analyzed Media Corpus dataset

- Unreliable media uses clickbaits more than mainstream media.
- Non-news type broadcast media practice more clickbait than news type broadcast media.

Media	Category	Clickbait	Non-clickbait	Clickbait (%)
Mainstream	Print	128022	402820	24.12
	Broadcast (news)	28543	99314	22.32
	Broadcast (non-news)	141209	87886	61.64
Unreliable	Clickbait	172271	203662	45.82
	Conspiracy	90389	224574	28.7
	Junk Science	23637	28935	44.96
	Satire	21798	19399	52.91

Amount of clickbaits in various media

Name	Category	Clickbait	Non-Clickbait	Clickbait(%)
VH1	Broadcast (non-news)	13760	1339	91.13
Amplifying Glass	Unreliable	692	71	90.69
MTV	Broadcast (non-news)	42313	4492	90.4
ClickHole	Unreliable	8250	930	89.87
Reductress	Unreliable	3984	484	89.17
Bravo TV	Broadcast (non-news)	8263	1242	86.93
Food Network	Broadcast (non-news)	2990	492	85.87
OWN	Broadcast (non-news)	474	118	80.07
E!	Broadcast (non-news)	24501	6167	79.89
Food Babe	Unreliable	2387	638	78.91

Top-10 clickbait practitioners

Acknowledgement

- Sabbir Md. Adnan Siddiqui - for his support in building the extension

References

- A. Chakraborty, B. Paranjape, S. Kakarla, and N. Ganguly. Stop clickbait: Detecting and preventing clickbaits in online news media. In Advances in Social Networks Analysis and Mining (ASONAM), 2016.
- A. Anand, T. Chakraborty, and N. Park. We used neural networks to detect clickbaits: You won't believe what happened next! ArXiv preprint arXiv:1612.01340, 2016.

BaitBuster in Action

