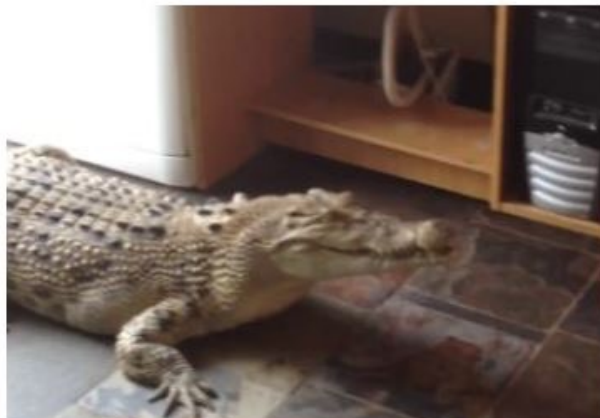


Stop Clickbait: Detecting and Preventing Clickbaits in Online News Media

Abhijnan Chakraborty, Bhargavi Paranjape,
Sourya Kakarla, Niloy Ganguly
IIT Kharagpur, India

You'll Get Chills When You See These Examples of Clickbait



**There Was A Crocodile Outside Her House, But
What She Did Next Is Just Insane**

You'll Get Chills When You See These Examples of Clickbait



**They Were Stuck In A Raging Flood, But What They
Did For Her Will Bring You To Tears**

You'll Get Chills When You See These Examples of Clickbait



**9 Serial Killers Who Were Never Caught – Holy Hell,
I'm Scared...**

What is Clickbait?

- (On the Internet) content whose main purpose is to attract attention and encourage visitors to click on a link to a particular web page. - Oxford English Dictionary
- Exploit **Curiosity Gap**:
 - Headlines provide forward referencing cues to generate painful **information gap**.
 - Readers feel compelled to click on the link to fill the gap, and ease the pain.

Good: Increased Viewership

TRENDING

282,603 VIEWS



Which Harry Potter House Does Your Cat Belong In?

Time to break out the old Sorting Hat and figure out exactly what kind of cats we have.



17 Pictures Of Bill Clinton Realizing That Balloons Exist

3,981,492 views

Good: Skyrocketing Valuations



BuzzFeed

BuzzFeed valued at more than
three times the Washington Post

Bad: RIP Journalistic Gatekeeping



Clickbait: The changing face of online journalism



PBS NEWSHOUR

MAKING SENSE

Column: Why click-bait will be the death of journalism



Is Clickbait the End of Online News?

How the race to the bottom dominates the world of online journalism

GOAL OF THIS WORK

Bring in more transparency and
Offer readers choice to deal with clickbaits

Workplan

- Given an article headline on a webpage, or on social media sites, detect the headline as clickbait, and warn the reader.
- Depending on reader choices, automatically block certain clickbait headlines from appearing on websites during her future visits.

How to Detect Clickbaits?

- Using fixed rules/matching common patterns: 74% accuracy
- URL/Domain name matching: not all stories of a domain are clickbaits (e.g., BuzzFeed news).

Detecting clickbaits is non-trivial!

To identify features, need to compare clickbaits
with traditional news headlines.

Dataset

Clickbait

- Collected 8,069 articles from BuzzFeed, Upworthy, ViralNova, Thatscoop, Scoopwhoop.
- 7,623 articles were annotated by volunteers as clickbaits.

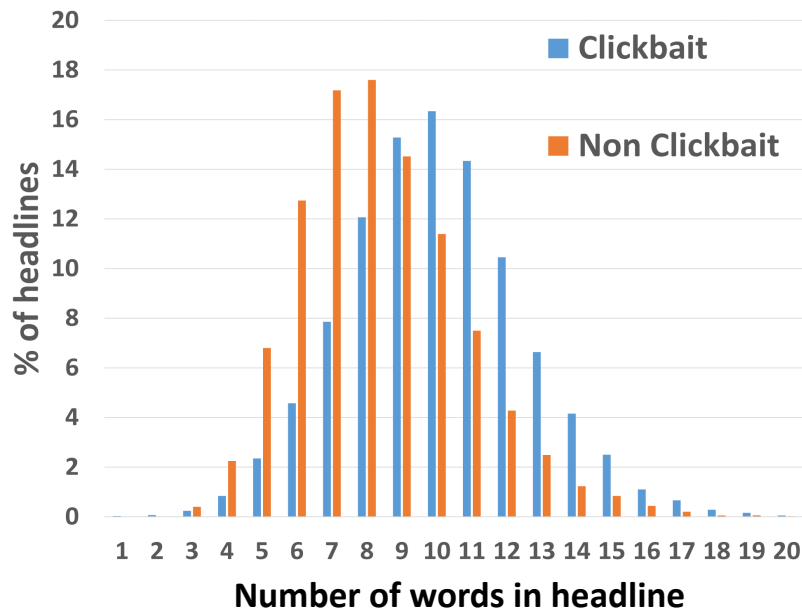
Non-clickbait

- Collected 18,513 articles from Wikinews.
- Community verified news content.
- Fixed guidelines to write headlines, rigorously checked.

Took 7,500 articles from each category for comparison.

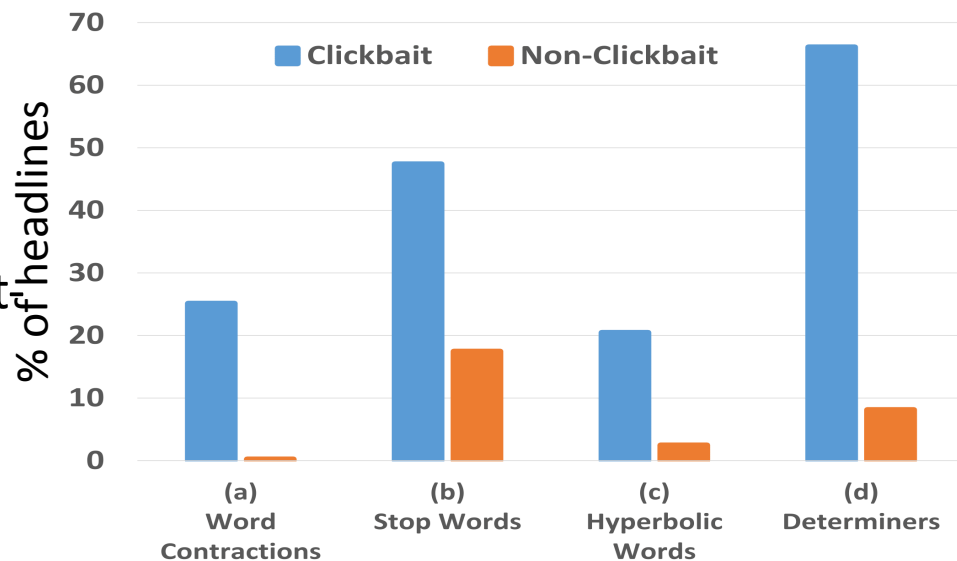
What makes clickbaits different?

- **Length:** Clickbaits are well formed English sentences that include both *content* and *function* words.
- **Unusual Punctuation Patterns:** Often ends with *!?*, *...*, *****, *!!!*
- **Use of Stop Words:** Disproportionate occurrence in clickbaits



What makes clickbaits different?

- **Word Contractions**: they're, you're, you'll, we'd
- **Words with very positive sentiment (Hyperbolic words)**: Awe-inspiring, breathtakingly, gut-wrenching, soul-stirring
- **Determiners** (forward reference particular people or things in the article): their, this, what, which

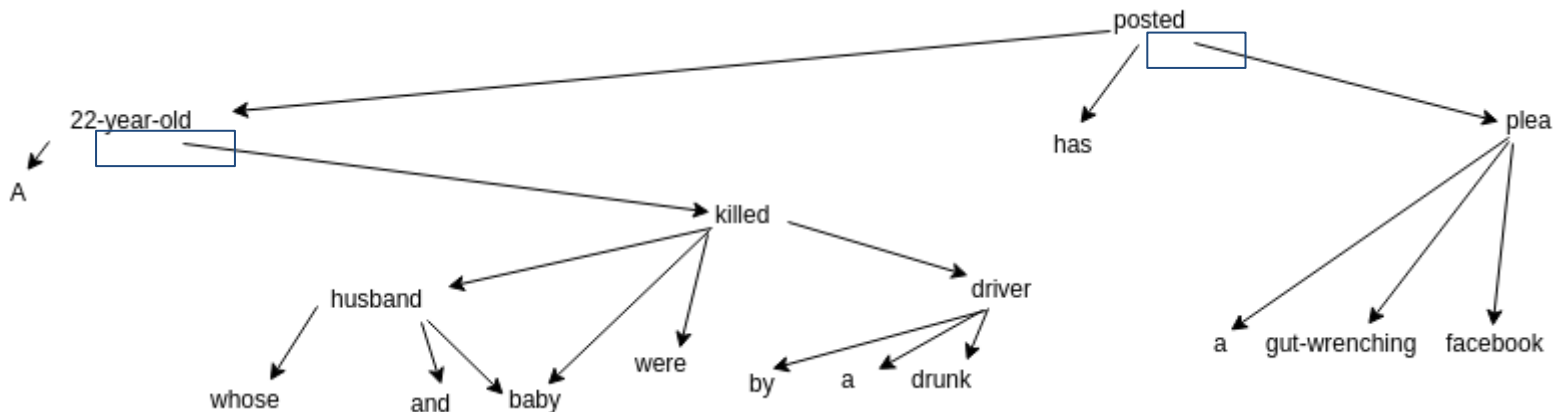


What makes clickbaits different?

Long Syntactic Dependencies between governing and dependent words:

- Due to existence of complex phrasal sentences.
- Length between subject '22-Year-Old' and verb 'Posted' is 11 in

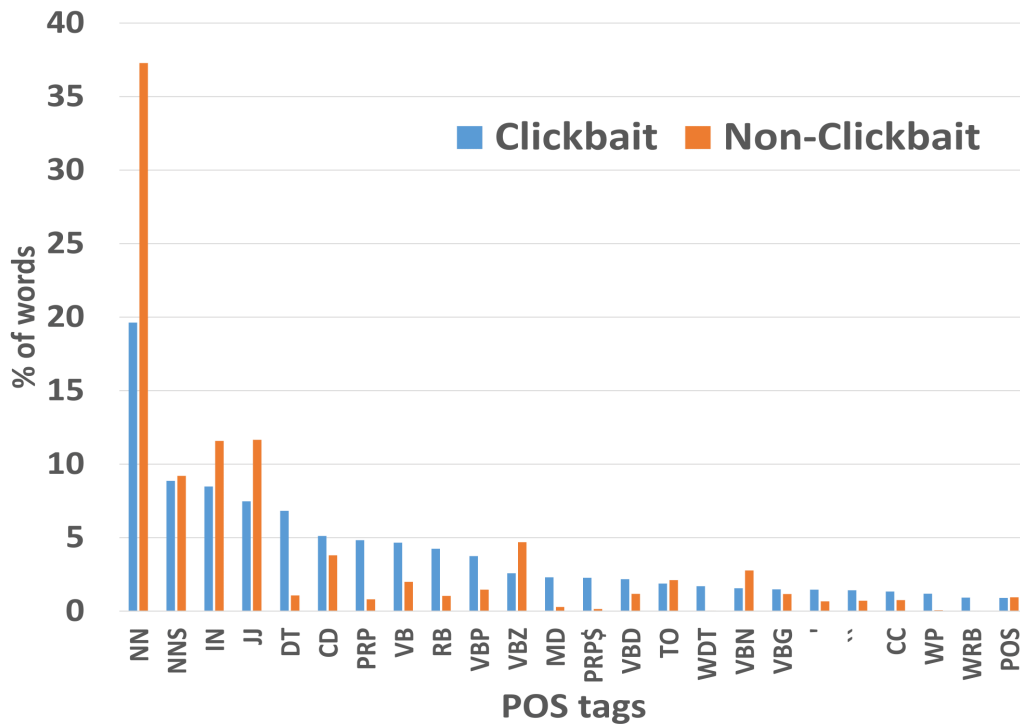
*A 22-Year-Old Whose Husband And Baby Were Killed By
A Drunk Driver Has Posted A Gut-Wrenching Facebook Plea*



What makes clickbaits different?

Distribution of POS tags

- **Non-clickbaits:** More proper nouns (NN), verbs in past participle and 3rd person singular form (VBN, VBZ).
- **Clickbaits:** More adverbs and determiners (RB, DT, WDT), personal and possessive pronouns (PRP, PRP\$), verbs in past tense and non-3rd person singular forms (VBD, VBP).



Classifying Headlines as Clickbaits

- Classifier: SVM with RBF kernel
- 14 Features (detailed in the paper).
- 10-fold cross validation performance:

Accuracy	93%
Precision	0.95
Recall	0.90
F1 Score	0.93

NEXT TASK

Block clickbaits from
appearing on different websites

What Interests You, Annoys Me

- 12 regular news readers reviewed 200 random clickbait headlines.
- Marked clickbaits they would click or block.
- Average Jaccard coefficients for clicked as well as blocked clickbaits are low across readers.
- Signals high heterogeneity in reader choices.

Reimagine Blocking as Personalized Classification!

Modeling Reader's Interests

- Model the reader's interests from the articles she has already clicked as well as already blocked.
- Two possible interpretations of reader interests in Clickbait (or lack thereof)
- For the following clickbait:
Can You Guess The Hogwarts House of These Harry Potter Characters?
 1. The reader likes/dislikes Harry Potter or the fantasy genre
 2. She likes/gets annoyed by the pattern, "Can You Guess"

Blocking Based on Topical Similarity

1. Extract content words from headline, article metatags and keywords that occur in the html <head>: **tagset**
2. Use BabelNet: multilingual semantic network which connects 14 million concepts and named entities.
3. Interest Expansion: Common hypernym neighbours of tags in tagset form a cluster (**nugget**). Two nuggets merge when nodes occur commonly in them.
4. Form reader's **BlockNuggets** and **ClickNuggets**.
5. Blocking decision on Query Tagset: How many nodes are common with BlockNuggets or ClickNuggets.

Blocking Based on Patterns

1. Normalization of headlines

- Numbers and Quotes are replaced by tags < D > and < QUOTE >
- 200 most common words + English stop words retained.
- Nouns, Adjectives, Adverbs and Verb inflections replaced by POS tags.

“Which Dead ‘Grey’s Anatomy’ Character Are You”

→ “which JJ < QUOTE > character are you”

→ “Which ‘Inside Amy Schumer’ Character Are You”

→ “which < QUOTE > character are you”

2. Set of patterns for both blocked and clicked articles.

3. Blocking decision on Query: Average word level Edit Distances from blocked and clicked articles.

Performance of Blocking Approaches

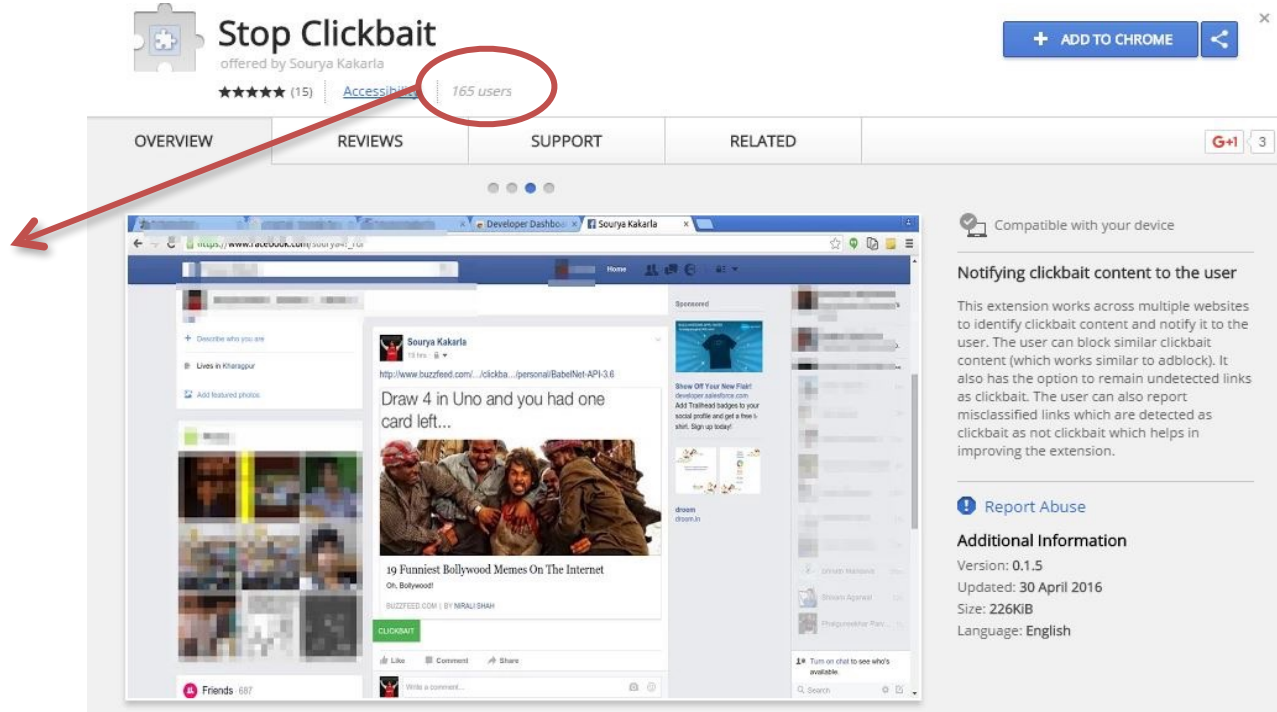
- 12 readers were shown 200 clickbait articles.
- Their blocks and clicks recorded.
- 3:1 train:test split with 4 fold cross validation.
- Pattern based approach performs best.

Approach	Accuracy	Precision	Recall	F1 Score
Pattern Based	0.81	0.834	0.76	0.79
Topic Based	0.75	0.769	0.72	0.74
Hybrid	0.72	0.766	0.682	0.72

Browser Extension: Stop Clickbait

Google chrome extension to report and block clickbaits on different websites.

165 chrome users have used it so far!



Notify Clickbaits

YES
Y'ALL



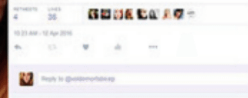
The 22 Most Disappointing Things That Have Ever Happened

CLICKBAIT

LOL



my favorite game is to act like i can't understand something very simple when a man is explaining it to me to see how dumb he thinks i am



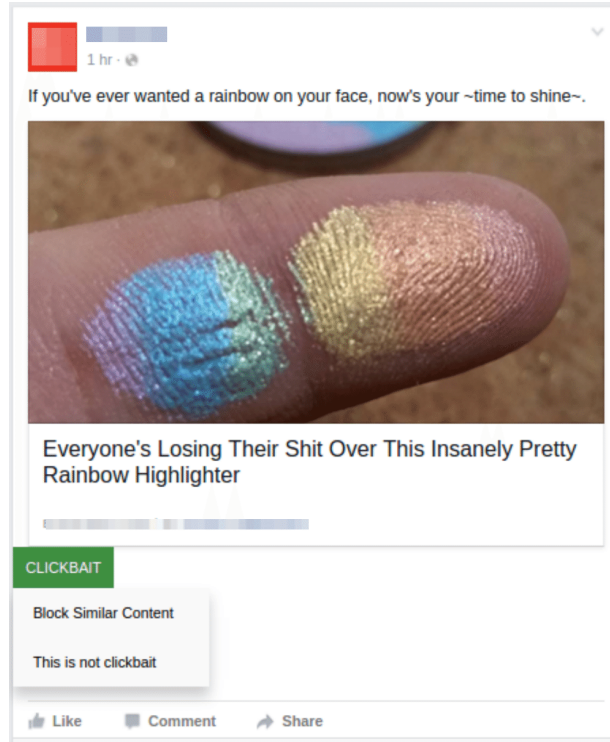
People Are Losing Their Minds Over How Hard This Woman Trolled A Random Dude

CLICKBAIT



5 days ago

Block or Report Wrong Label



Report Missed Clickbaits

TWEETS
21.9K

FOLLOWING
1,841

FOLLOWERS
519K

LIKES
2,352

LISTS
6

 Upworthy @Upworthy · 22h

21 interesting facts about penguins that may just put a smile on your face.
#WorldPenguinDay u.pw/1QvGQc8



Retweet 39

Like 31

More options

Open link in new tab

Open link in new window

Open link in incognito window

Save link as...

Copy link address

 Report Clickbait

Inspect Ctrl+Shift+I

*Thank
you*

