

Product Spam on YouTube: A Case Study



Janek Bevendorff



Matti Wiegmann



Martin Potthast



Benno Stein

Leipzig University, Bauhaus-Universität Weimar
Intelligent Information Systems Group – webis.de

OSSYM 2023 – October 04–06, CERN, Geneva, Switzerland



🔍 best baby jumpers



🔍 best baby jumpers

🔍 best baby jumpers **for hips**

🔍 best baby jumpers **and bouncers**

🔍 best baby jumpers **uk**

🔍 best baby jumpers **and activity centers**

🔍 best baby jumpers **for infants**



best baby jumpers and swings



Shopping

Images

Videos

News

Maps

Books

Flights

Finance

About 25.400.000 results (0,49 seconds)



YouTube

[https://www.youtube.com > watch](https://www.youtube.com/watch)

Baby Bouncers, Rockers, Swings & Jumperoos - YouTube



Baby Bouncers, Rockers, Swings & Jumperoos - WHICH ONE IS BEST? · Comments.

YouTube · The Mom and Baby House · 17 Nov 2022



YouTube

[https://www.youtube.com > watch](https://www.youtube.com/watch)

Best Baby Jumper - Top Baby Jumparoo Reviews - YouTube



Buy **Best Baby Jumpers/Jumparoo** By just Clicking These "Amazon Affiliate Links". ✓1. Bright Starts Door Jumper: ◦Amazon US: ...

YouTube · For Sweet Littles · 24 Feb 2020



YouTube

[https://www.youtube.com > watch](https://www.youtube.com/watch)

Best Baby Door Jumpers - Gives Your Little One A Soft Place ...

[[YouTube](#)]



Q best blood glucose meters|



Q best blood glucose meters

Q best blood glucose meters **2023**

Q best blood glucose meters **2022**

Q best blood glucose meters **uk**

Q best blood glucose meters **app**

Q best blood glucose **meter in india**

Q best blood glucose **meter malaysia**



best blood glucose meters



Videos

Images

News

Books

Maps

Flights

Finance

About 15.500.000 results (0,50 seconds)



YouTube

<https://www.youtube.com/watch>

Best Blood Glucose Meters in 2022 [Top 5 Picks] - YouTube



Visit the link for the updated price: <https://theconfusionreduce.com/blood-glucose-meters> In This video, We recommended the top 5 best ...

YouTube · Confusion Reduce · 11 May 2022



4 key moments in this video



YouTube

<https://www.youtube.com/watch>

Top 5 Best Glucose Meters Review in 2023 - YouTube



best Glucose Meter featured in this video- 0:27 NO.1. Care Touch Diabetes Testing Kit – <https://linkwhirl.com/wxqYit> 1:09 NO.2.

YouTube · Trendy Ideas · 24 Apr 2020



5 key moments in this video

[YouTube](#)

[Want more?](#)

[Trouble sleeping?](#)

Why Do These Videos Exist?

Why Do These Videos Exist?

Two main explanations:


- ❑ Cheap web shop advertising.
- ❑ Click farming from affiliate links in the video description.

Why Do These Videos Exist?

Two main explanations:

- ❑ Cheap web shop advertising.
- ❑ Click farming from affiliate links in the video description.

Easy Money Affiliate Marketing Scheme:

1. Content creator (affiliate) registers with affiliate partner (usually a seller or manufacturer).
2. Affiliate partner issues tagged product referral links.
3. Affiliate spams users with these affiliate links.
4. Users click on links (and potentially buy products).
5. Affiliate gets  .

Can We Detect This Spam Automatically?

Can We Detect This Spam Automatically?

Well, let's see! We...

- ...crafted 7,392 product review queries using two public product taxonomies:¹
 `best <product category>`,
- ...queried Startpage (Google), Bing, and DuckDuckGo, and
- ...extracted all YouTube URLs from the SERPs.

¹GS1 Global Product Classification (GPC), Google Product Taxonomy (GPT)

YouTube Affiliate Spam Dataset

	Startpage	Bing	DDG	All
Total results	147,658	147,592	143,823	439,073
Unique URLs	128,854	122,775	112,702	258,400
Unique websites	41,514	26,853	22,862	60,947
YouTube URLs	5,033	1,127	847	7,007
Unique YouTube URLs	4,588	1,098	810	5,902
Transcripts available	4,242	785	426	4,993
Transcripts				4,755
Test data (ground truth obtained via manual labeling)				200
Training data (unsupervised labeling through clustering)				4,555

YouTube Affiliate Spam Dataset

	Startpage	Bing	DDG	All
Total results	147,658	147,592	143,823	439,073
Unique URLs	128,854	122,775	112,702	258,400
Unique websites	41,514	26,853	22,862	60,947
YouTube URLs	5,033	1,127	847	7,007
Unique YouTube URLs	4,588	1,098	810	5,902
Transcripts available	4,242	785	426	4,993
Transcripts				4,755
Test data (ground truth obtained via manual labeling)				200
Training data (unsupervised labeling through clustering)				4,555

YouTube Affiliate Spam Dataset

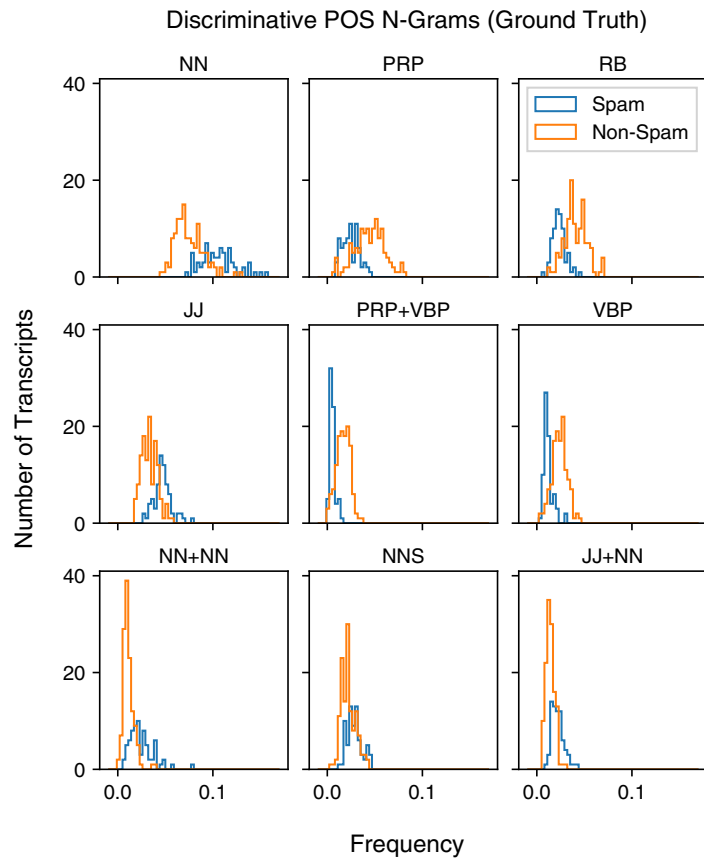
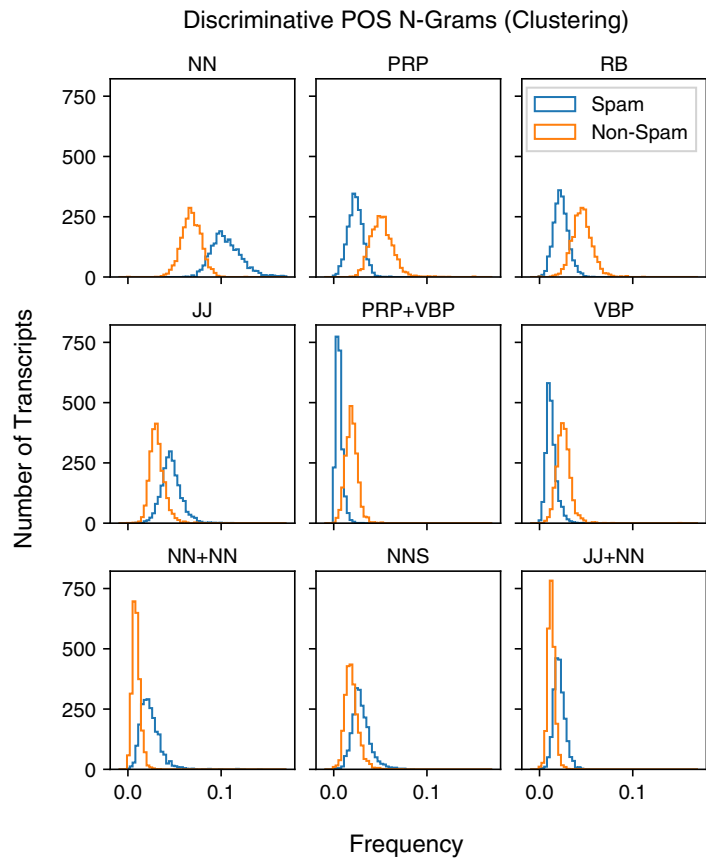
	Startpage	Bing	DDG	All
Total results	147,658	147,592	143,823	439,073
Unique URLs	128,854	122,775	112,702	258,400
Unique websites	41,514	26,853	22,862	60,947
YouTube URLs	5,033	1,127	847	7,007
Unique YouTube URLs	4,588	1,098	810	5,902
Transcripts available	4,242	785	426	4,993
Transcripts				4,755
Test data (ground truth obtained via manual labeling)				200
Training data (unsupervised labeling through clustering)				4,555

Features

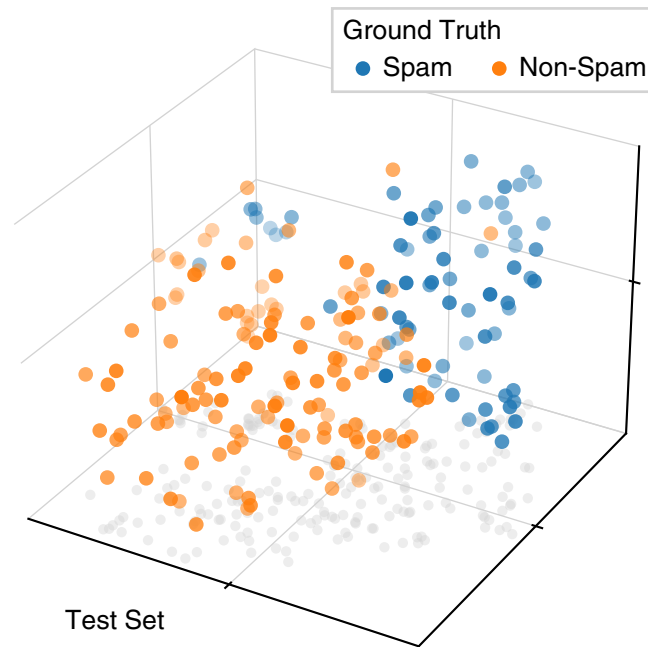
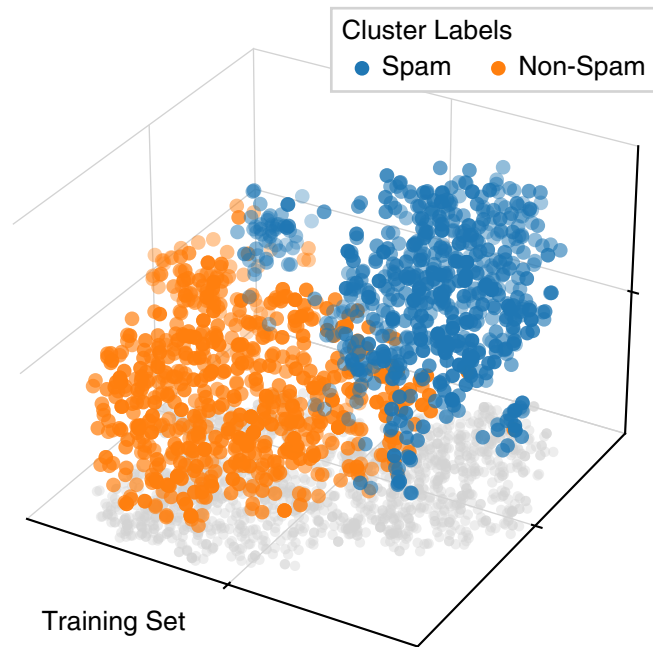
Part-of-speech (POS) n-grams ($n \in \{1, 2, 3, 4\}$) as topic-independent document representation.

Features

Part-of-speech (POS) n-grams ($n \in \{1, 2, 3, 4\}$) as topic-independent document representation.



Spectral Clustering as Silver Standard Training Data



Supervised Classification

Last Step: Learn cluster labels to produce a reusable model:

Model	Class	Prec.	Recall	F1	AUROC
SVM	Non-Spam	0.98	0.78	0.87	0.94
	Spam	0.70	0.97	0.82	
Log. Regression	Non-Spam	0.97	0.81	0.88	0.93
	Spam	0.73	0.96	0.83	

Supervised Classification

Last Step: Learn cluster labels to produce a reusable model:

Model	Class	Prec.	Recall	F1	AUROC
SVM	Non-Spam	0.98	0.78	0.87	0.94
	Spam	0.70	0.97	0.82	
Log. Regression	Non-Spam	0.97	0.81	0.88	0.93
	Spam	0.73	0.96	0.83	

More in our paper, all code and data publicly available:



github.com/webis-de/OSSYM-23

webis.de/publications