

Bogofilter

- Paul Graham, *A Plan for Spam*
- Bag of words
- Bayes theorem

<http://www.paulgraham.com/spam.html> Not Bayesian. Lexical filter. Supervised.

- Want: map from word to Pr email containing it is spam
- Bias to avoid false positives
- Minimum occurrence threshold
- Top 15 spam and ham words

Good weighs twice bad. At least 5 times to count (3 in spam due to doubling).

1 Break

Break

Questions?

purple.com/talk-feedback