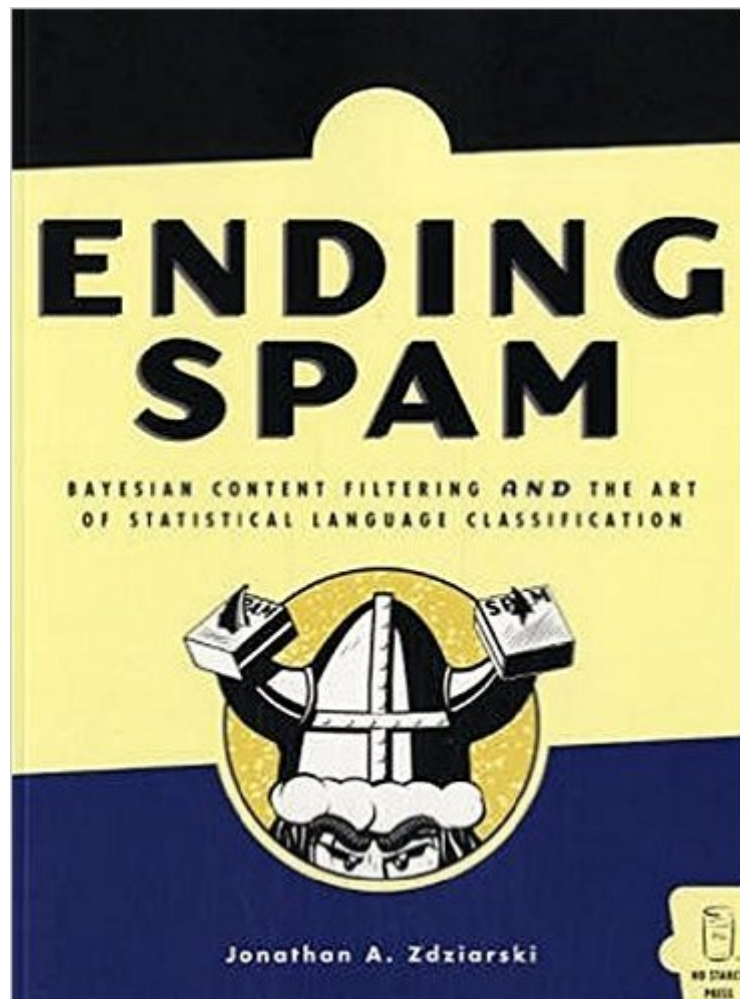# Ending Spam: Bayesian Content Filtering And The Art Of Statistical Language Classification

Synopsis

Join author John Zdziarski for a look inside the brilliant minds that have conceived clever new ways to fight spam in all its nefarious forms. This landmark title describes, in-depth, how statistical filtering is being used by next-generation spam filters to identify and filter unwanted messages, how spam filtering works and how language classification and machine learning combine to produce remarkably accurate spam filters.After reading Ending Spam, you'll have a complete understanding of the mathematical approaches used by today's spam filters as well as decoding, tokenization, various algorithms (including Bayesian analysis and Markovian discrimination) and the benefits of using open-source solutions to end spam. Zdziarski interviewed creators of many of the best spam filters and has included their insights in this revealing examination of the anti-spam crusade. If you're a programmer designing a new spam filter, a network admin implementing a spam-filtering solution, or just someone who's curious about how spam filters work and the tactics spammers use to evade them, Ending Spam will serve as an informative analysis of the war against spammers.TOCIntroductionPART I: An Introduction to Spam FilteringChapter 1: The History of SpamChapter 2: Historical Approaches to Fighting SpamChapter 3: Language Classification ConceptsChapter 4: Statistical Filtering FundamentalsPART II: Fundamentals of Statistical FilteringChapter 5: Decoding: Uncombobulating MessagesChapter 6: Tokenization: The Building Blocks of SpamChapter 7: The Low-Down Dirty Tricks of SpammersChapter 8: Data Storage for a Zillion RecordsChapter 9: Scaling in Large EnvironmentsPART III: Advanced Concepts of Statistical FilteringChapter 10: Testing TheoryChapter 11: Concept Identification: Advanced TokenizationChapter 12: Fifth-Order Markovian DiscriminationChapter 13: Intelligent Feature Set ReductionChapter 14: Collaborative AlgorithmsAppendix: Shining Examples of FilteringIndex

Book Information

Paperback: 312 pages

Publisher: No Starch Press; 1 edition (July 1, 2005)

Language: English

ISBN-10: 1593270526

ISBN-13: 978-1593270520

Product Dimensions: 7 x 0.7 x 9.2 inches

Shipping Weight: 1.1 pounds (View shipping rates and policies)

Average Customer Review: 4.3 out of 5 stars See all reviews (14 customer reviews)

Best Sellers Rank: #2,184,821 in Books (See Top 100 in Books)   #75 in Books > Computers &

Customer Reviews

Ending Spam from Mr. Zdziarski is a well written BASIC and easy to understand INTRODUCTION to get a technical overview of todays spam fighting solutions on the market.Also it is written on the cover that it is f.e focused towards developers, network admins etc. I would consider the target customer to be IT Managers, or other curious people who want to get an overview.Thats what it does and it does it very well in my eyes.The book provides simplified, abstract overviews of some available spam filters solutions.The book is provided into 3 parts- An Introduction part to spam filtering (Chapter 1-4)- A part describing "Fundamentals of Statistical Filtering" (Chapter 5-9)- an the third part describing "Advanced Concepts of Statistical Filtering" (Chapter 10-14)Its a bit confusing that Chapter 4 has the same title than Part II. So perhaps Chapter 4 should have been part of "Part II" ?The Chapters which I found most interesting were:Chapter 4 "Fundamentals of Statistical Filtering"Chapter 7 "The Low down dirty Tricks of spammers"Chapter 9 "Scaling in Large Environments"I am sure the author could have easily filled the book with Chapter 7 alone. The book is very entertaining and has a nice motivating writing style. You might at times find some rant about the spammers which I have chosen to ignore as it doesnt contain any valuable information or anything which I didnt know already. While I might agree to some of the authors views, I believe that the rant does unfortunately do exactly the opposite in my eyes and does give spammers credit to how they do their work.

Download to continue reading...

Ending Spam: Bayesian Content Filtering and the Art of Statistical Language Classification Bayesian Signal Processing: Classical, Modern and Particle Filtering Methods (Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control) Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference (Addison-Wesley Data & Analytics) Body Language: Body Language Training - Attract Women & Command Respect, by Mastering Your High Status Body Language (Body Language Attraction, Body Language ... Language Secrets, Nonverbal Communication) WHO Classification of Tumours of the Urinary System and Male Genital Organs (IARC WHO Classification of Tumours) WHO Classification of Tumours of Soft Tissue and Bone (IARC WHO Classification of Tumours) DDC 22 Dewey Decimal Classification and Relative Index (Dewey Decimal Classification & Relative Index) Introduction to

Cataloging and Classification, 10th Edition (Introduction to Cataloging & Classification (Paperback)) Classification System for Church Libraries: Based on the Dewy Decimal Classification System HOW TO MAKE MONEY OPPOSING SPAMMERS - If You receive SPAM You can turn it into PROFIT Absolute Beginner's Guide to Security, Spam, Spyware & Viruses Elementary Stochastic Calculus With Finance in View (Advanced Series on Statistical Science & Applied Probability, Vol 6) (Advanced Series on Statistical Science and Applied Probability) Thermodynamics With Quantum Statistical Illustrations. Monographs in Statistical Physics and Thermodynamics, Volume 2 FrameMaker - Creating and Publishing Content: LEARN TO USE, MANAGE, AND PUBLISH CONTENT WITH ADOBE FRAMEMAKER BODY LANGUAGE : Decoding Alpha Male Body Language, Instantly Attract Any Woman Without Saying a Single word. (Body Language 101, Alpha male, Attract woman, ... Seduce Women, Eye Contact, Body Language) Content Everywhere: Strategy and Structure for Future-Ready Content Content is King: How to use great SEO content, video and analytics to put you ahead of the game Documents, Presentations, and Workbooks: Using Microsoft Office to Create Content That Gets Noticed- Creating Powerful Content with Microsoft Office Creating Fat Content: Boost Website Traffic with Visitor-Grabbing, Google-Loving Web Content Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory