

Desmascarando mensagens apócrifas

José Paulo de Araújo – UFRJ

Internet e Comunicação

- ▶ Popularização a partir de 1996
- ▶ Serviço mais popular: e-mail
 - Agilidade
 - Armazenamento
 - Assincronicidade
 - Replicabilidade



Spam

- ▶ Mensagens não-solicitadas
- ▶ 80% das mensagens circulantes
- ▶ Problemas:
 - Provedores/Empresas
 - Aumentam custos de serviços
 - Armazenamento
 - Banda
 - Filtragem
 - Usuários
 - Aumentam tempo para seleção
 - Queda de produtividade



Spam

- Impacto ambiental
 - Emissões de CO₂ equivalentes a...

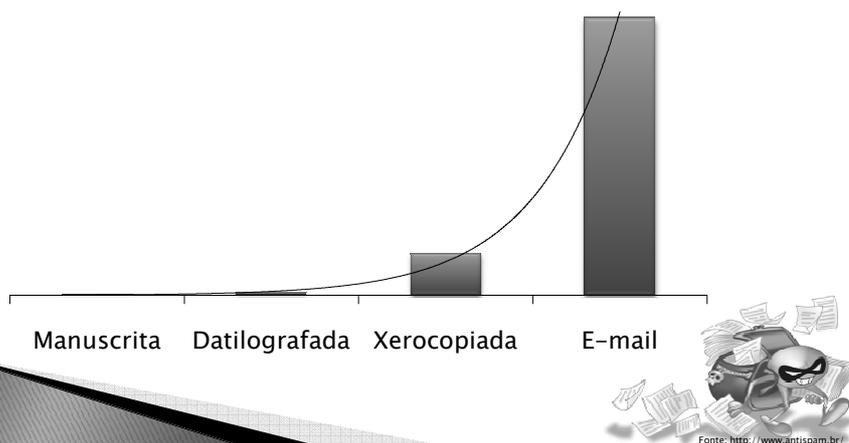
3 milhões de carros!

Fonte: McAfee e ICF



✉ Correntes

► Propagação:



✉ Correntes por e-mail

CATEGORIA	SUBCATEGORIA	EXEMPLO ¹³	
CORRENTE	-	Carta-corrente ^(S) , Corrente anticorrente ^(M)	
	Saúde/Doença Sortel/Azar	Carta-corrente tradicional ^(C) , Correntes de Sta. Edwírges ^(M) , Corrente de ameaça ^(C) , Paródia ^(M) , Corrente-teste ^(M)	
	Humor	Piada ^(C) , Piada ^(M) , Trote ^(M)	
	Amor/amizade	Amor e amizade ^(M)	
	GANHO/ PERDA	Humanitarismo	Corrente com fim real ^(M) , Ativismo de escritório ^(M)
		Solidariedade Apoio	Pedido de ajuda ou solidariedade ^(C) , Criança doente ^(M) , Pelas crianças ^(M)
	Oferta	Brinde ^(C) , Dinheiro ^(M) , Algo por nada ^(M)	
	Fraude	Corrente de fraude ^(C) , Fraude ^(S)	
VERACIDADE	-	Lenda real ^(C) , Verdade ^(C) , Real, mas... ^(M) , Meia-verdade/meia-mentira ^(C) , Imagem imperfeita ^(M) , História alterada ^(C) , Não é piada ^(M)	
ALERTA	Virus	Alerta de virus ^(M) , Informe sobre virus ou trojans ^(M) , Alerta de código malicioso ^(C) , Alarme falso ^(S) , Mal-entendido ^(S) , Boato de virus ^(S) , Falso virus ^(C) , Alarme ^(S)	
	Mito/lenda	Corrente alarmista ^(C) , Para o bem de sua saúde ^(M) , Mito urbano ^(C) , Alerta inócuo ^(C)	
	Contextual	América sob ataque ^(S)	

Quadro 9 – Categorias de Correntes Digitais

ARAÚJO, J.P. Correntes digitais: evolução e classificação. Disponível em: <www.comunicar.pro.br/artigos/hoaxes_araujo_2003.pdf> 2003.



✉ Apócrifas

- ▶ Correntes que reivindicam autoria de autores conhecidos:
 - Millôr Fernandes, Luiz Fernando Verissimo, Arnaldo Jabor, João Ubaldo Ribeiro, Caetano Veloso, Jorge Luis Borges, Carlos Drummond de Andrade, Gabriel García Márquez etc.
- ▶ Problemas para o suposto autor:
 - Quebra de direito autoral alheio
 - Conteúdo ofensivo atribuído a ele



✉ Jabor e os apócrifos

Já reclamei aqui desses textos apócrifos, mas tenho de me repetir. Não dá mais. Todo dia surge na internet uma nova besteira, com dezenas de emails me elogiando pelo que eu não fiz. [...] o pior são artigos escritos em meu nome por inimigos covardes para me sujar.



✉ Como saber?

Tenho horror a mulher perfeitinha. Sabe aquele tipo que faz escova toda manhã, tá sempre na moda e é tão sorridente que parece garota-propaganda de processo de clareamento dentário?

A verdade é que nossa lei é um labirinto sem saída, um buraco que ninguém sabe como encher. Dessa confusão surgem os quebra-galhos: chamem o Exército! Os bombeiros! As ONGs! As milícias mineiras...



✉ Atribuição de autoria

- ▶ Atividade da Linguística Forense
- ▶ Idioletos ou estilos pessoais
 - Ortografia: erros recorrentes
 - Seleção vocabular
 - Padrões sintáticos recorrentes
 - Pontuação
- ▶ Princípios:
 - Não há semelhanças absolutas entre escritores
 - Ninguém escreve da mesma forma todo o tempo



✉ Metodologia

- ▶ Textos de Arnaldo Jabor
 - O Estado de São Paulo (Amostra A: 18)
 - Jornal da Globo (Amostra B: 30)
- ▶ Textos apócrifos (Amostra C: 11)
 - Web
 - Cora Rónai
- ▶ Uso do MS Word®
 - Ortografia e Gramática
 - Estatísticas de Legibilidade
 - Palavras/Sentença
 - Caracteres/Palavra
 - Escore Flesch



✉ Escore Flesch

- ▶ Facilidade/dificuldade de leitura
- ▶ $206,835 - (1,015 \times MPS) - (0,846 \times MSP)$
 - MPS = Média de Palavras por Sentença
 - MSP = Média de Sílabas por Palavra



Escore	Legibilidade	Nível de Instrução
100-75	Muito Fácil	1° ao 4° do EF
75-50	Fácil	5° ao 8° do EF
49-25	Difícil	EM e Superior
20-0	Muito Difícil	Academia

✉ Resultado da análise

- ▶ Autênticos: legibilidade alta
- ▶ Apócrifos: legibilidade muito alta
- ▶ Variação entre A e B
 - ▶ Significância estatística?



Escore Flesch	Amostra A	Amostra B	Amostra C
Muito Fácil	6%	27%	73%
Fácil	94%	73%	27%
Difícil	-	-	-
Muito Difícil	-	-	-

✉ Testes estatísticos

- ▶ Amostras
 - ▶ não-pareadas
 - ▶ Independentes
- ▶ Teste t bicaudal
- ▶ Teste não-paramétrico de Mann-Whitney



	A x C	B x C
Fator	Significância	Significância
Palavras	Muito Alta	Muito Alta
Palavras/Sentença	Alta	Não
Caracteres/Palavra	Alta	Não
Escore Flesch	Muito Alta	Sim

Discussão

- ▶ Procedimento com potencial
- ▶ Necessidade de análise adicional:
 - Qualitativa
 - Quantitativa
- ▶ Mais estudos
- ▶ Conscientização do usuário
 - Não acreditar
 - Não encaminhar
 - Espalhar (não por e-mail!) as dicas acima



antispam.br

Fonte: <http://www.antispam.br/>

