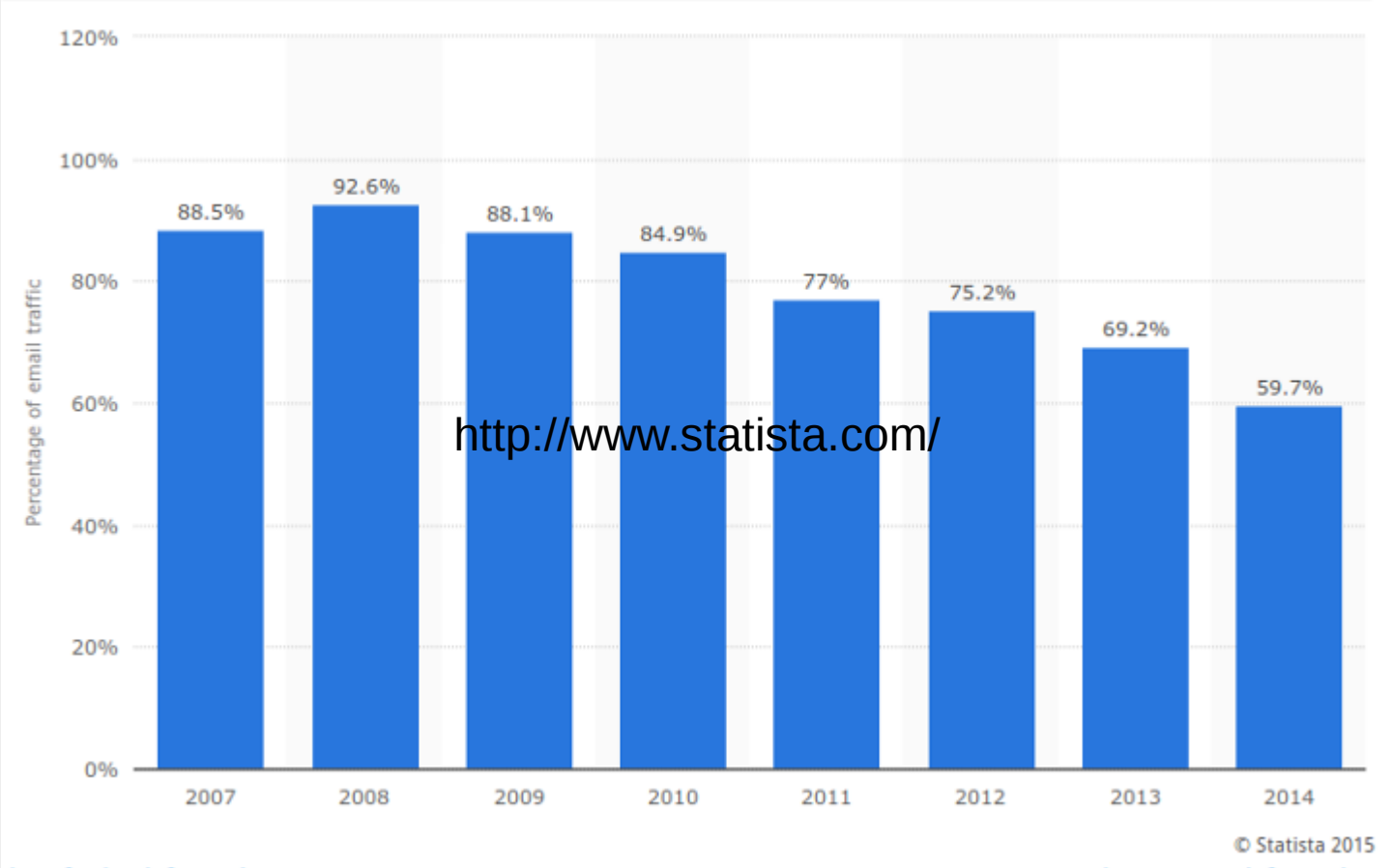
Что такое спам и как с ним бороться (если вы лингвист)

Мария Рубинштейн, лингвист,
старший спам-аналитик
“Лаборатории Касперского”,
Москва

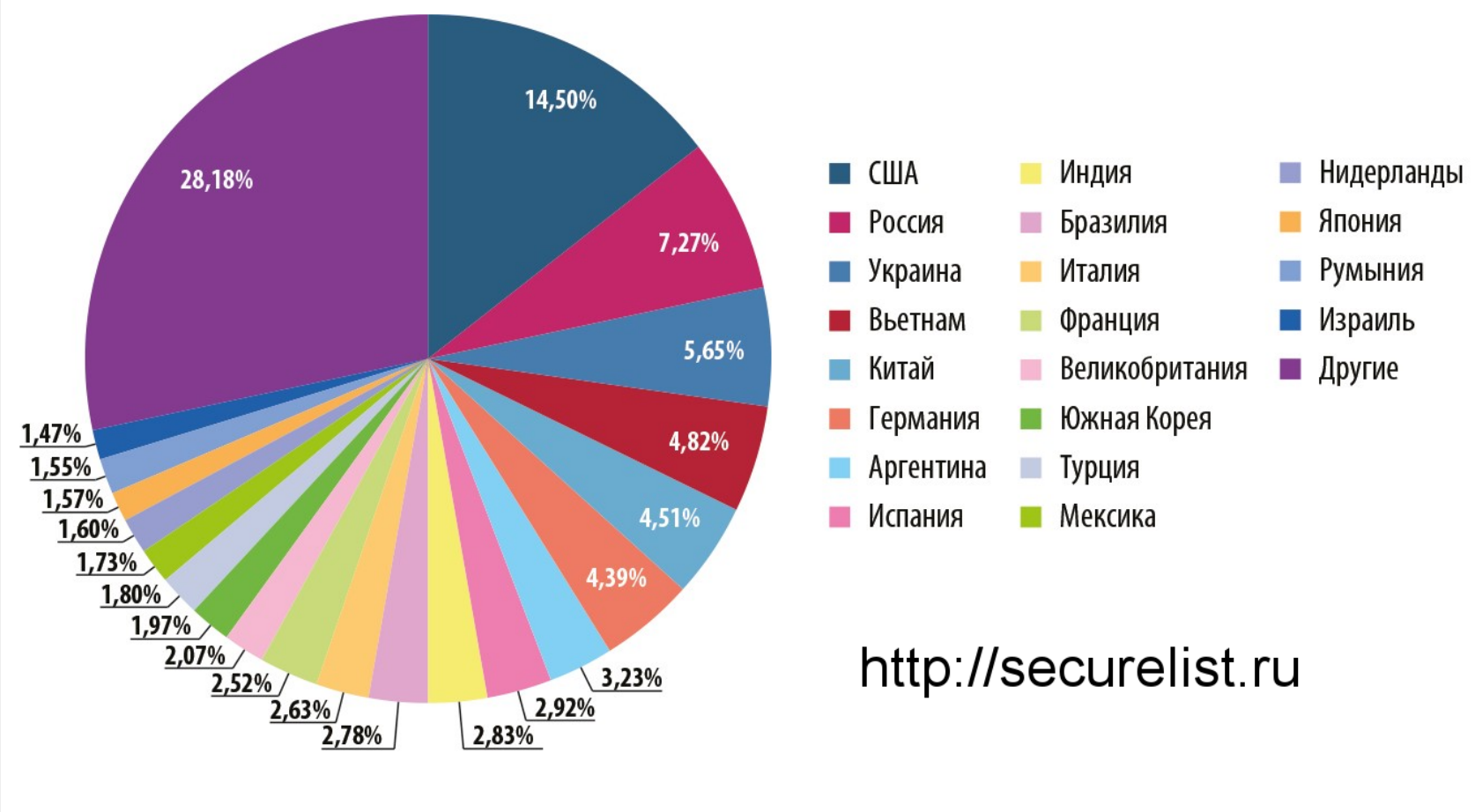
Что такое спам

- ***Спам:*** массовая анонимная незапрошенная почтовая рассылка
- SPiced hAM
- Первая спамерская рассылка: 1978 (ARPANET); первый коммерческий спам: 1994

Доля спама в мировом почтовом трафике



Кто рассылает спам и зачем



Кто рассылает спам и зачем

Разбираемся в сортах спама

- Реклама
  - Средства от облысения дешево!
- Рассылка вирусов и прочих вредоносных программ
  - Уважаемый клиент, в архиве уведомление о платеже.
- (Компьютерное) мошенничество
  - Поздравляем! Вы выиграли миллион в лотерею!
  - Ваш банковский счет заблокирован, пройдите по ссылке и введите номер карточки и пин-код.
- PR и “черный PR”
  - Голосуйте за нашу партию!
  - Наша компания оказывает криминальные услуги!
- Тестирование спамерских баз
  - АУщацукп цукпщцру уканг. щжлхшщ

Задачи спамеров

Письмо должно:

- прийти в почтовый ящик адресата
- быть понятным адресату
- обойти спам-фильтры

Для адресата

- Письма составляются на разных языках (обычно с помощью Google Translate)
- Письма делаются короткими
- Письма содержат завлекательный текст (*Dear Sir/Madam, you won $10000 in the Coca Cola lottery, to claim your prize, contact Mr ...*)
- Письма маскируются под личные обращения (*Привет, как ты? Все еще ходишь в офис? Я открыл новый способ заработать, зацени!*)

Против спам-фильтров

- Текст-картинка
- Письмо с зашумлением текста
- Гиперссылка на страницу в интернете
- Текст письма в приложенном файле
- Маскировка под личную переписку
- Иносказания, использование синонимов

Задачи спам-фильтра

Контентная фильтрация: анализ текста письма

Фильтр распознает спам:

- по наличию определенных слов и выражений, их сочетаний
  - спамеры владеют разными языками и Google Translate'ом
  - спамеры не всегда грамотны, а иногда специально зашумляют текст
- по структуре письма (например, одна гиперссылка и больше ничего)
- по зашумлению текста и другим искажениям

Примеры спама: письма на разных языках

- Dobrý deň, priateľ. Som Dr Christopher Johnson vedúca odboru účtovníctva Audit Nat West banka, Harlesden, North West London, tu v Anglicku (NatWest Bank) tu v Anglicku. Píšem vám o obchodnej návrh, ktorý bude mať obrovský prínos pre nás oboch. V mojom oddelení, je vedúci London Krajský úrad, som zistil, súčet £ 16,5 milióna (šestnásť Milión a pol milióna libier) v účte, ktorý Patrí k jednému z našich zahraničných zákazníkov Neskoré Obchodné Mogul pán Moises Saba Masri Miliardár, Žid z Mexika, ktorý bol obeťou zrútenie vrtuľníka 10.1.2010, zabíjať ho a jeho rodinných príslušníkov. Saba bolo 46-rokov-starý. Aj v vrtuľníka v čase havárie bolo jeho manželka, ich syn Abraham (Alberto) a jeho dcéra-in-law. Pilot bol tiež mŕtvy.
- Hallo Freund. Ich bin Dr. Christopher Johnson Head of Accounting Audit Department eines Nat West Bank, Harlesden, North West London, hier in England (Natwest Bank) hier in England. Ich schreibe Ihnen über ein Unternehmen Vorschlag, der eine immense Vorteil für uns beide sein wird. In meiner Abteilung, wobei die Manager London Regional Office, entdeckte ich eine Summe von £ 16,5 Millionen (sechzehn Millionen und 500.000 Pfund Sterling) in einem Konto, das zu einem unserer ausländischen Kunden Late Business-Mogul Mr. Moises Saba Masri Billionaire gehört, ein Jude aus Mexiko, die ein Opfer von einem Hubschrauberabsturz 10. Januar 2010 war, ihn zu töten und seine
- Hey hälsningar, Jag är Mr Arthur Ryan chef för Kassör kommitten en bank från Harlesden, nordvästra London, här i England (NatWest Bank London). Jag skriver dig om en affärside som kommer att vara en enorm fördel för oss båda. I min avdelning, chef Greater London Regional Office, jag upptäckte en summa av £ 16.500.000 (Sexton miljoner femhundra tusen pund) på ett konto som tillhör en av våra utländska kunder Late Business Mogul Mr Moises Saba Masri miljardär, en Judisk från Mexiko som var ett offer för en helikopterkrasch i början av förra 2 år, dödade honom och familjemedlemmar. Saba var 46-år gammal. Även i helikoptern vid tidpunkten för olyckan, hans hustru, deras son Avraham (Albert) och hans dotter-in-law. Piloten var också död.

Примеры спама: письма на разных языках

From: Rosina JillianTagro <rosineahie69@msn.com>

To: undisclosed recipients:

Bonan tagon,

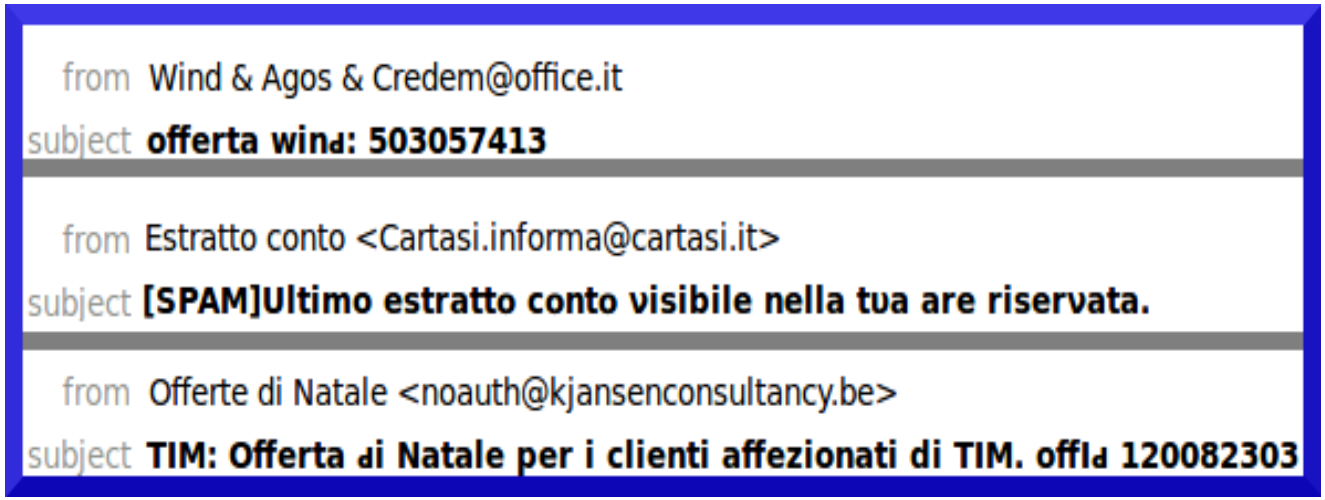
Mia nomo estas Rosina Jillian Tagro la filino de deziro Asségnini Tagro. Mia patro estis Ivorian politikisto kiu servis kiel ministro pri internaj kaj stabestro de kaj enerala sekretario de la elpelita iama Ivorian Prezidanto Laurent Gbagbo dum la 2002-2012 Ivorian politika krizo / civila milito kaj li estis anka supro aliancano de la elpelita eksa prezidanto Laurent Gbagbo. Estas mal oja diri ke li mortis mardon 12an de aprilo 2011 en PISAM hospitalo post li estis kelke batante kaj fermis la prezidanta restadejo de armitaj ribeluloj / respublikaj fortoj de Eburbordo (FRCI) lojala al Alassane Ouattara Dramane kiu estas nun la prezidanto de mia lando Eburbordo. i tie vi povas vidi la video de mia patro kopion al PISAM hospitalo en Abidjan, kie li fine mortis http://www.youtube.com/watch?v=DDJhv3hps1g

Mia patrino Hawa inkludante miaj fratoj Zika, Goba, Bouabre kaj mia pli juna fratino Fatim estis iuj mortigitaj en Deukoué vila o dum ili fu is de la politika krizo / civila milito kaj mi estis en lernejo en Ganao dum la politika krizo / civila milito en Eburbordo tial mi estas la sola postvivanto en mia familio. Jen la video de la masakro en Déukoue; http://www..youtube.com/watch?v=kCJO9-y2P6Y Anta la morto de mia patro en PISAM hospitalo en Cocody Abidjan, li informis min pri sia ku ejo valora Du milionoj okcent mil e roj nur (2,800,000.00 €), kiun li deponis kun mia nomo kiel la sola heredanto al la fundo. Nun mi estas denove en Eburbordo kaj mi konfirmis la ekziston de la mono en la banko kaj la tuta konfirmite kovrante la mono estas nerompita.

Bonvolu Mi bezonas vian specon kaj ur a helpo al kopii kaj renversi tiun monon en vian landon kaj anka veni en vian landon por da rigi mian edukaj portanto. Kiam mi ricevos vian ur a respondo indikante vian intereson helpi min sukcese trapasi la monon por via lando kaj venu super al via lando, mi donos al vi iujn necesajn informojn vi povas postuli al procedi al trapasante la monon. Fine, mi pretas oferti al vi 20% de la tuta mono modo de kompenso pro via penado por helpi min per i tiu humila peto ar mi kredas ke tiu transakcio estus finis ene de malmultaj tagoj vi signifas via intereso por helpi min.

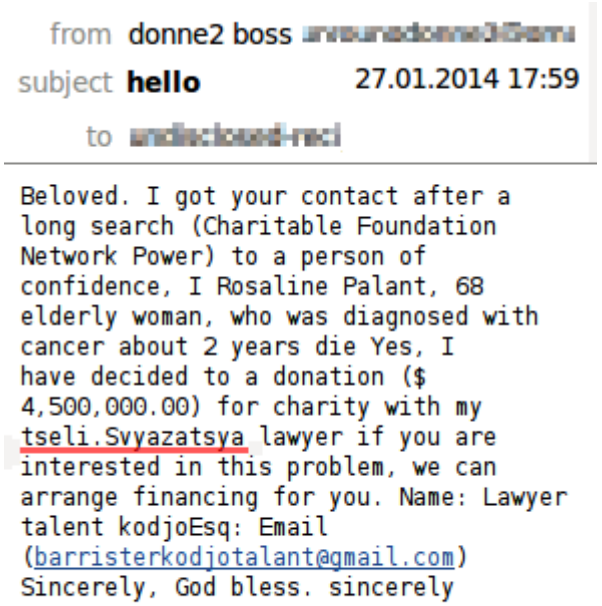
Dankon kaj Dio benu vin. Via tenere Rosina Jillian Tagro.

Примеры спама: одни символы вместо других



Примеры спама: неудачный перевод

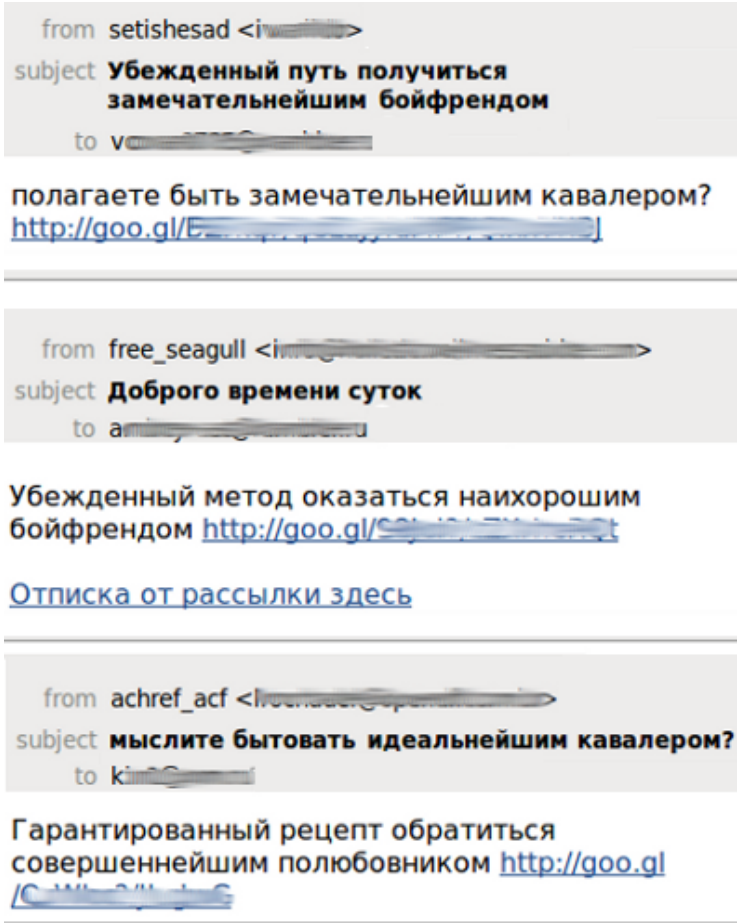
“Нигерийское письмо” (мошенничество)



from donne2 boss [redacted]
subject **hello** 27.01.2014 17:59
to [redacted]

Beloved. I got your contact after a long search (Charitable Foundation Network Power) to a person of confidence, I Rosaline Palant, 68 elderly woman, who was diagnosed with cancer about 2 years die Yes, I have decided to a donation ($ 4,500,000.00) for charity with my tseli.Svyazatsya lawyer if you are interested in this problem, we can arrange financing for you. Name: Lawyer talent kodjoEsq: Email (barristerkodjotalant@gmail.com) Sincerely, God bless. sincerely

Примеры спама: синонимы



from setishesad <i
subject **Убежденный путь получиться замечательнейшим бойфрендом**
to v

полагаете быть замечательнейшим кавалером?
http://goo.gl/E

---

from free_seagull <i
subject **Доброго времени суток**
to a

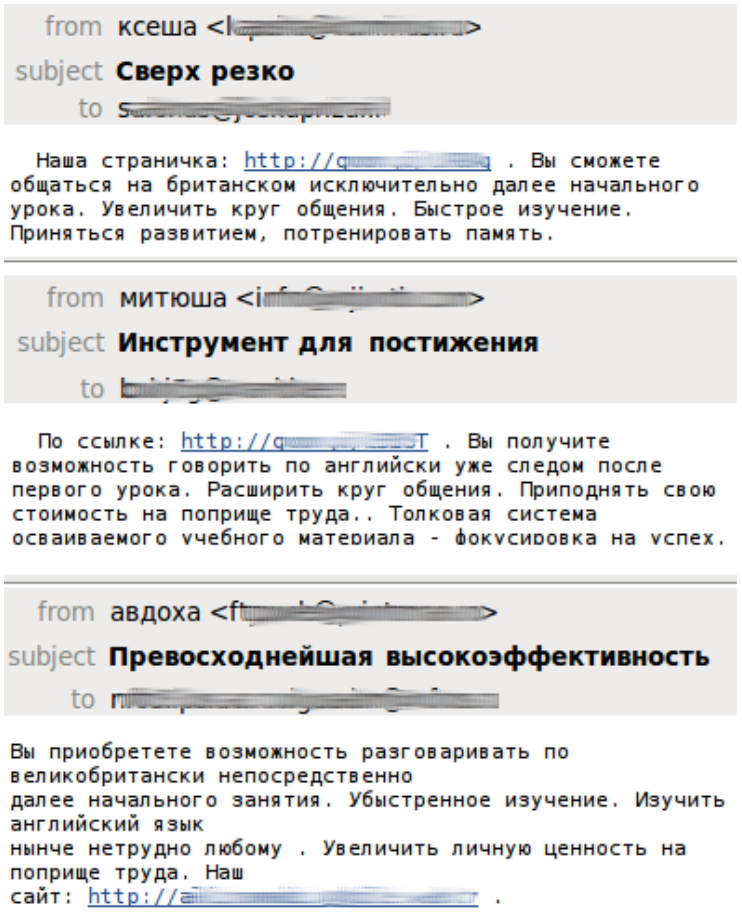
Убежденный метод оказаться наихорошим бойфрендом http://goo.gl/

Отписка от рассылки здесь

---

from achref_acf <
subject **мыслите бытовать идеальнейшим кавалером?**
to k

Гарантированный рецепт обратиться совершеннейшим полюбовником http://goo.gl
/



from ксеша <
subject **Сверх резко**
to

Наша страничка: http://q . Вы сможете общаться на британском исключительно далее начального урока. Увеличить круг общения. Быстрое изучение. Приняться развитием, потренировать память.

---

from митюша <i
subject **Инструмент для постижения**
to

По ссылке: http://q T . Вы получите возможность говорить по английски уже следом после первого урока. Расширить круг общения. Приподнять свою стоимость на поприще труда.. Толковая система осваиваемого учебного материала - фокусировка на успех.

---

from авдоха <f
subject **Превосходнейшая высокоэффективность**
to

Вы приобретете возможность разговаривать по великобритански непосредственно
далее начального занятия. Убыстренное изучение. Изучить английский язык
нынче нетрудно любому . Увеличить личную ценность на поприще труда. Наш
сайт: http://a .

Примеры спама: синонимы

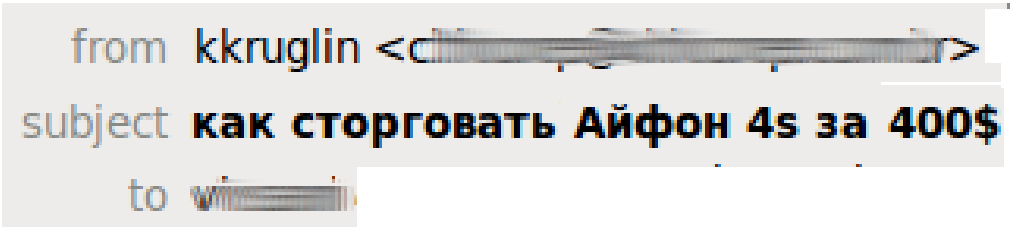
From Михаил <rt@[redacted]>

Subject **комм. предложение: сбыт новых беларусских тракторов в сборе с последней модели малогабаритными кабинами** 13.03.2014 22:07

To info@[redacted]

```
Приветствуем! Предлагаем предложить|Рекомендуем|Представляем}
известнейшие о-пропашные|} общего назначения беларусские трактор{а|}
МТЗ-82{.1|-80.1|} о смонтированными|оборудованные|уснащенными|снабженными}
современными  малыми кабинами|с современными  кабинами малого размера}
{{российского изготовления|производства РФ|}||}|разработки и сборки} в
России|собственной разработки и сборки}

Наш сайт http://[redacted]
```



Вы отроду не задумывались над тем, что расценки в магазинах вполне могут иметься больно прибавлены? http://[redacted]

Отписка от рассылки здесь

Примеры спама: искажение текста

Китайский:

вместо

税 [shuì] `получение налога'

употребляют два иероглифа

禾兑 [hé duì] `изюм обмен' (бессмысленное словосочетание)

...и как с ним бороться? (особенно если вы лингвист)

Контентная фильтрация - один из методов борьбы со спамом

- Распознавание зашумленного текста, искаженных написаний
  - buuuuuuy vÏåğraaaaaaa > buy viagra
- Анализ текста (с учетом морфологии)
- Поиск заведомо “спамерских” слов и их сочетаний (для разных языков)
  - `наследство' + `000 долларов' + `мой покойный клиент' + `риска нет' + `напишите нам' = *нигерийское письмо*
- Определение тематики письма (ключевые слова и словосочетания) + анализ технических заголовков
  - Тематика: “Банковское уведомление”, заголовки: фальшивые
  - Тематика: “Эротика”, есть гиперссылки