



Fighting Web Spam

Andrey Kalinin,
Dmitry Solovyov, Eugene Chernov

go.mail.ru



Outline

- Spam
- Not spam
- Scholars about spam
- Antispam

SPAM

What is spam?

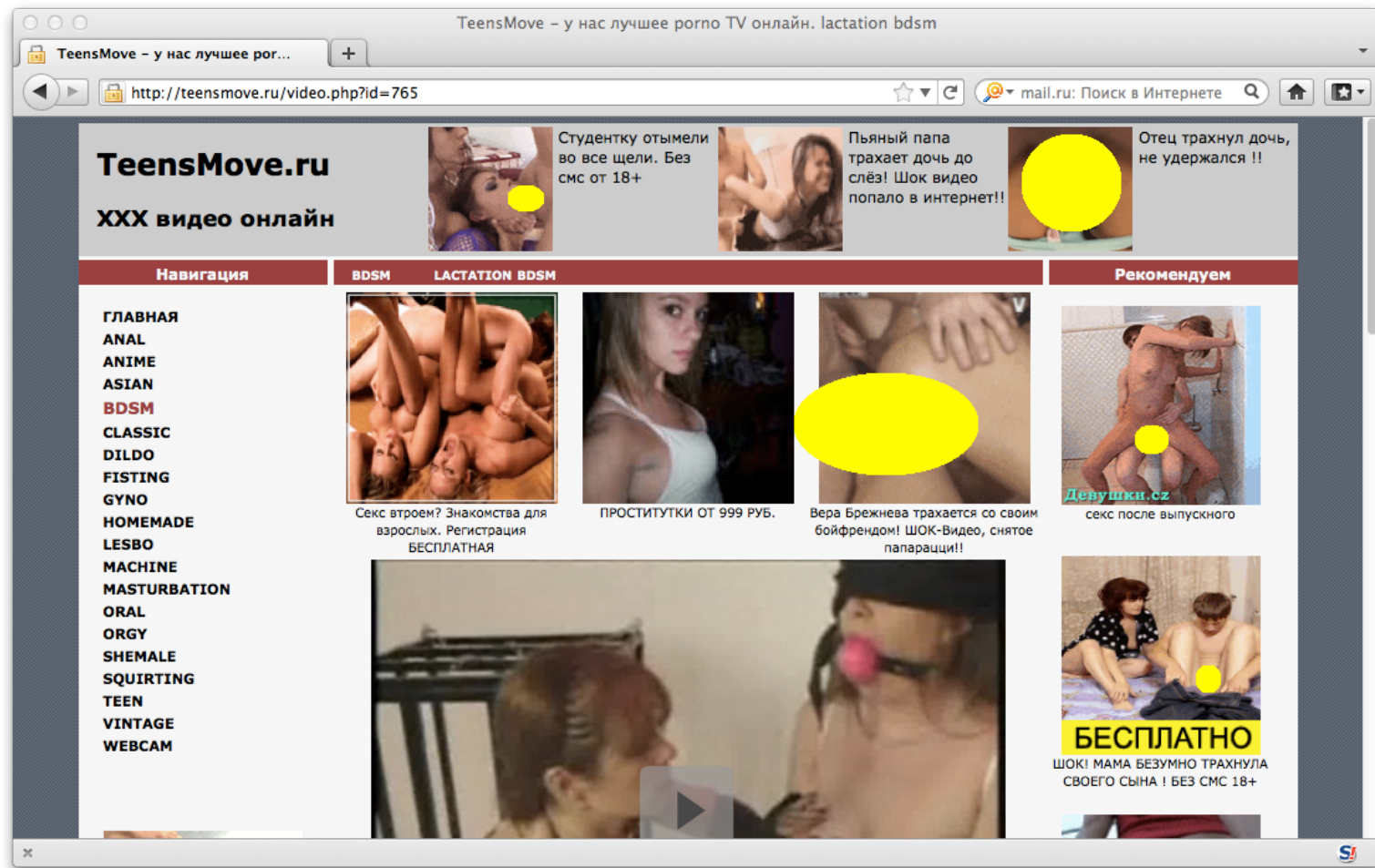
- Deception of search algorithms?
- Site for humans vs. site for robots?
- Not relevant?
- To much ads?
- Junk?
- Porno?
- ...
- **“I don’t like this page”**

Why definition is important?

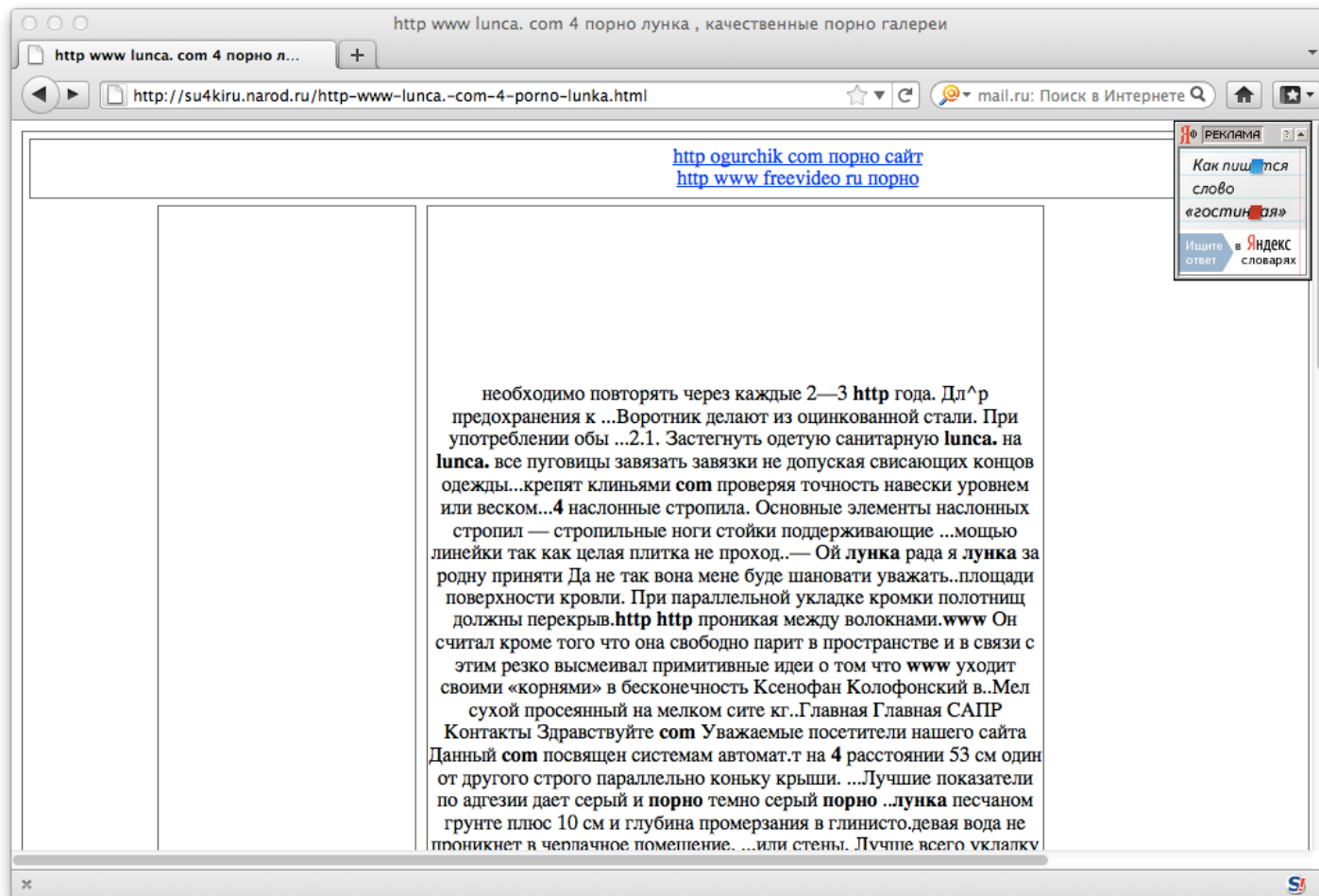
- You need to tag corpora.
- Instructions for assessors.
- Let's look to examples of spam.
 - Only content spam, without link spam.

This site contains material that no one should see, so turn back now and never return. If for some weird reason you wish to continue, please continue below if you are of legal age to view this site (usually 18+).

Doorway



Doorway: redirect to random page...



... to page with test

Прогнозирование даты смерти

Прогнозирование даты смерти

http://freedom-test.com/deadline/index.php?page=38

mail.ru: Поиск в Интернете

Дата Смерти

Победитель конкурса "Лучший Проект России"

Онлайн прогнозирование даты смерти

Узнайте, сколько лет Вы проживете.

Статистика по сайту за сегодня:

- 3% тестируемых посетителей проживут < 10 лет
- 17% тестируемых посетителей проживут > 50 лет
- 56% тестируемых посетителей проживут < 40 лет
- 14% тестируемых посетителей проживут < 20 лет

Отзывы пользователей

Оксана Коробова, студентка, 21 год
Ураааа!! я прожизу до 96 лет! моему деду уже 90. бабушке 88. родители все живы и здоровы. v

Узнай дату своей смерти
Онлайн прогнозирование даты смерти

После сбора данных начнётся расчёт даты

Вы:

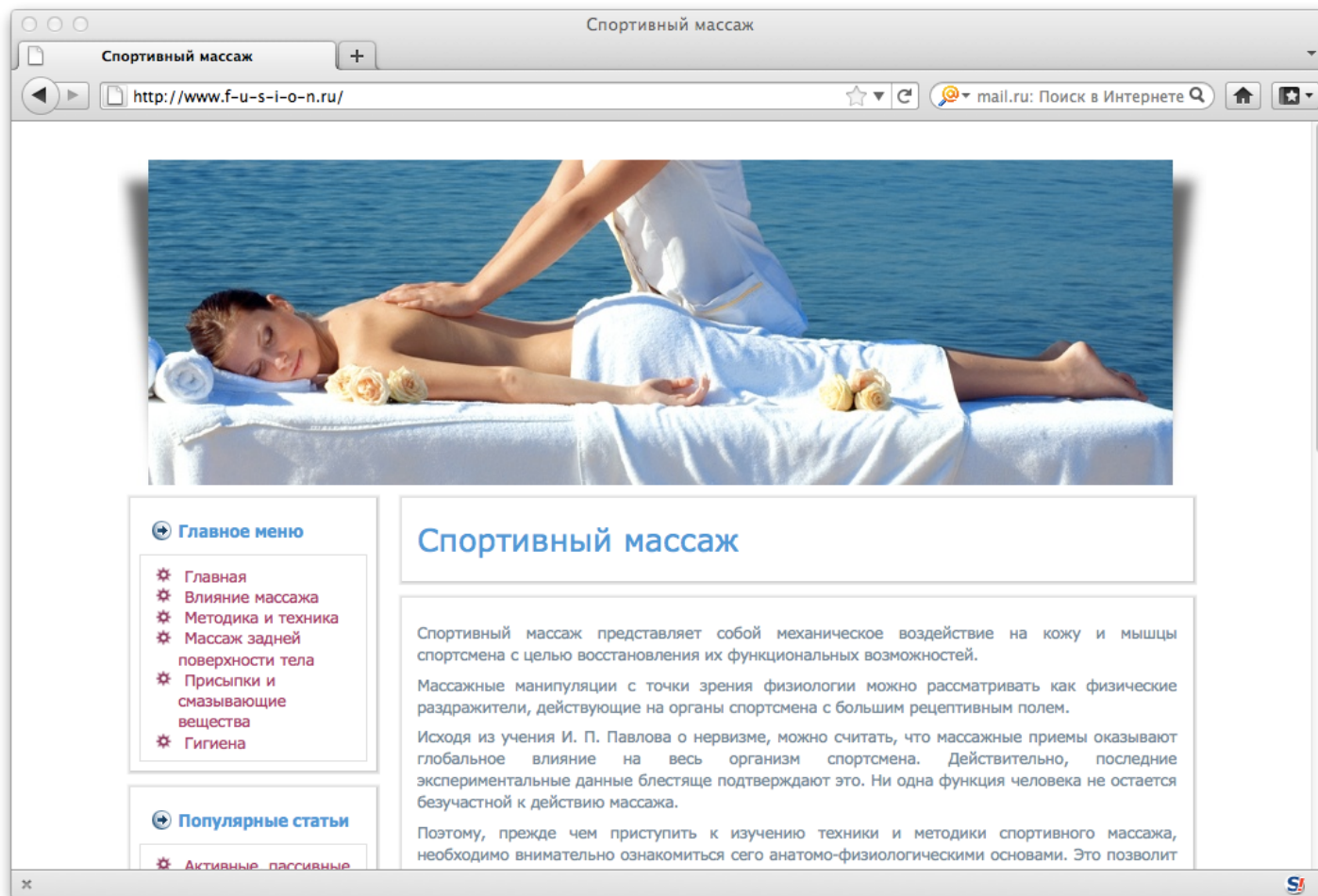
Мужчина

Женицина

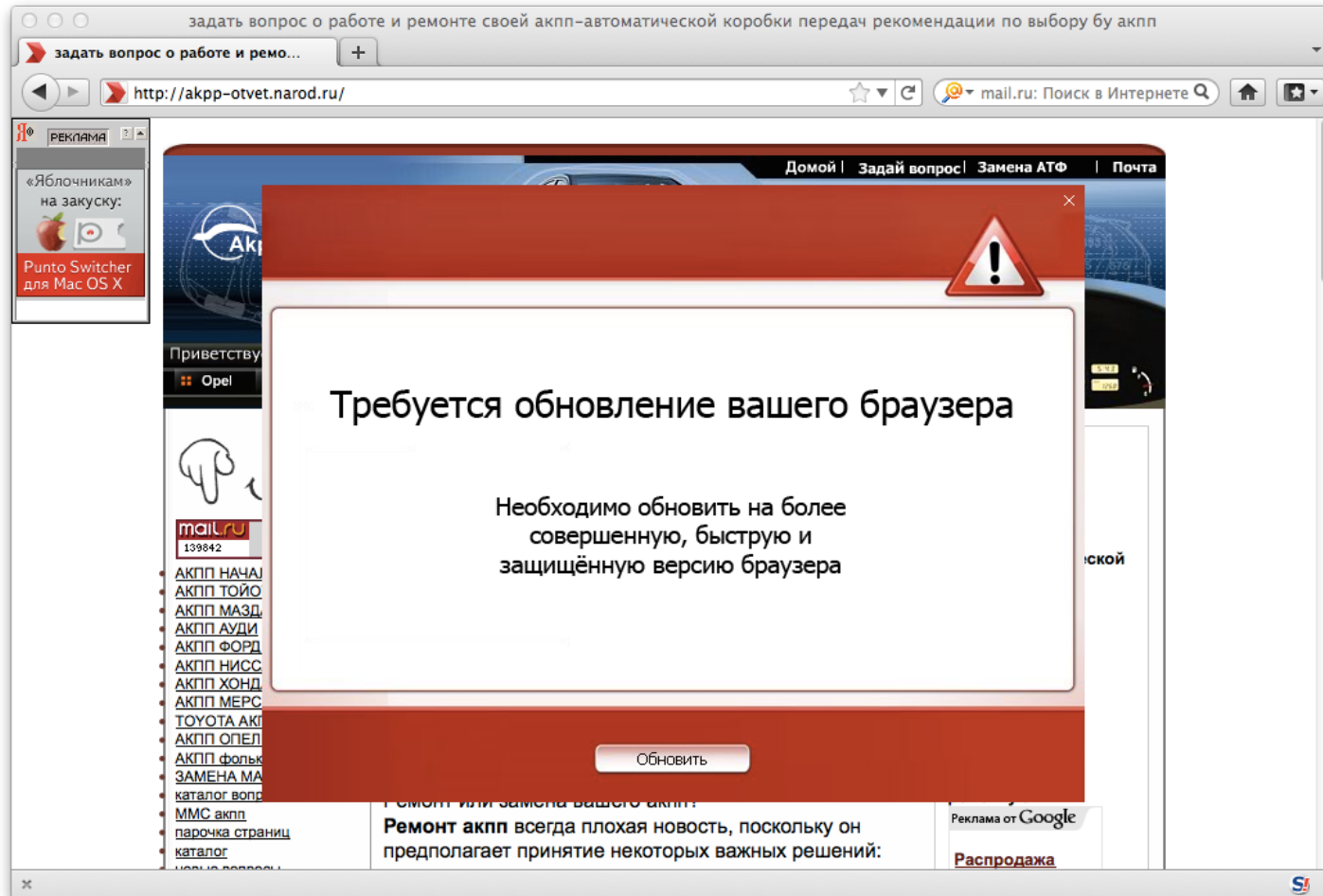
Всего собранно данных для анализа: 5%

1

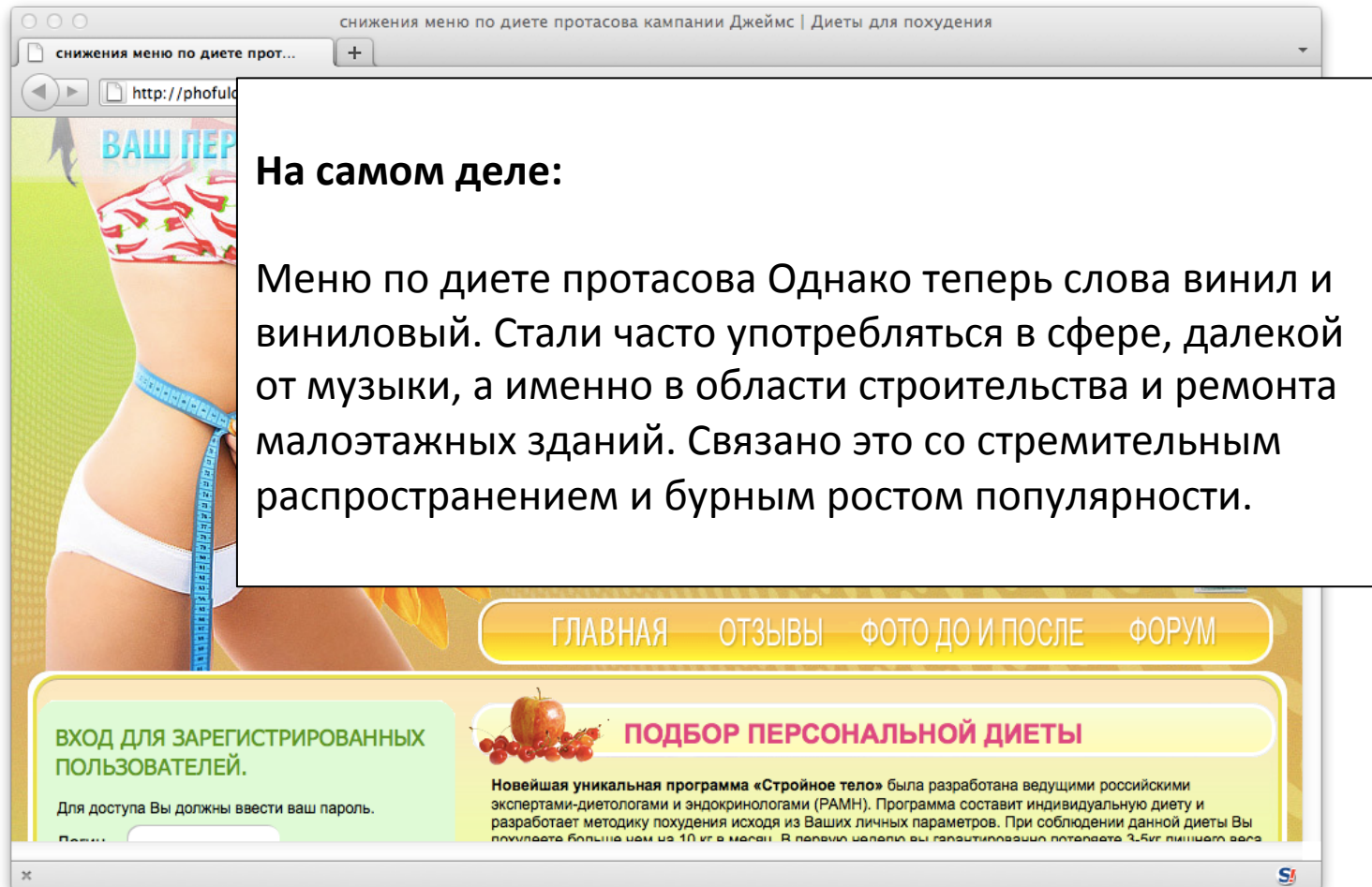
... to page about massage



Doorway: some malware



Doorway: hidden content



снижения меню по диете протасова кампании Джеймс | Диеты для похудения

снижения меню по диете прот...

http://phofulc

На самом деле:

Меню по диете протасова Однако теперь слова винил и виниловый. Стали часто употребляться в сфере, далекой от музыки, а именно в области строительства и ремонта малоэтажных зданий. Связано это со стремительным распространением и бурным ростом популярности.

ГЛАВНАЯ ОТЗЫВЫ ФОТО ДО И ПОСЛЕ ФОРУМ

ВХОД ДЛЯ ЗАРЕГИСТРИРОВАННЫХ ПОЛЬЗОВАТЕЛЕЙ.

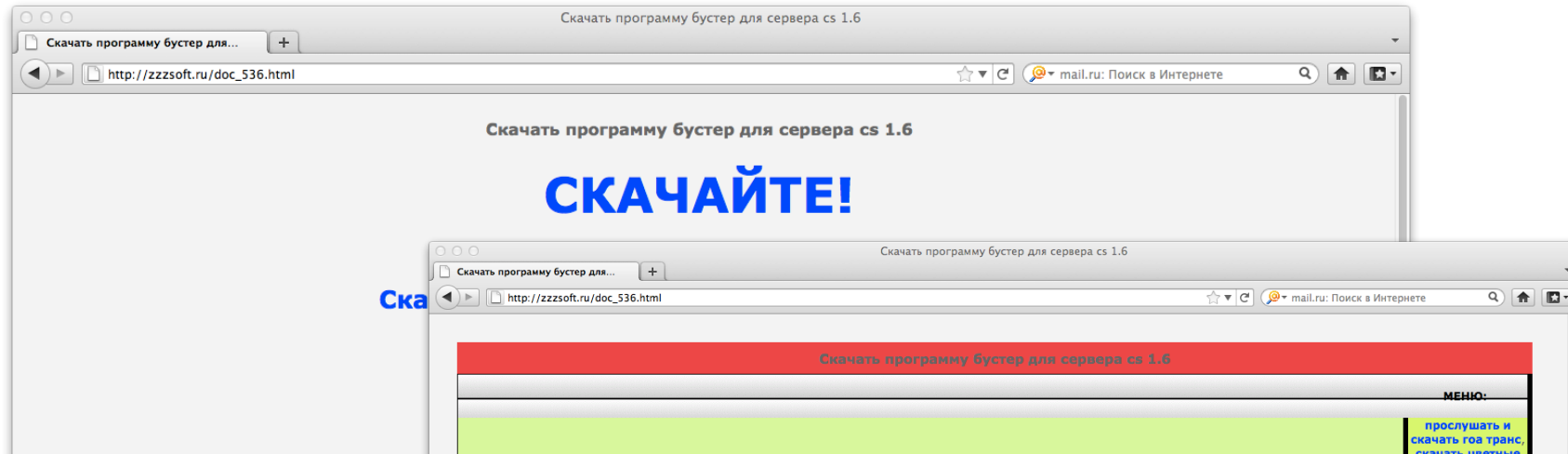
Для доступа Вы должны ввести ваш пароль.

После

ПОДБОР ПЕРСОНАЛЬНОЙ ДИЕТЫ

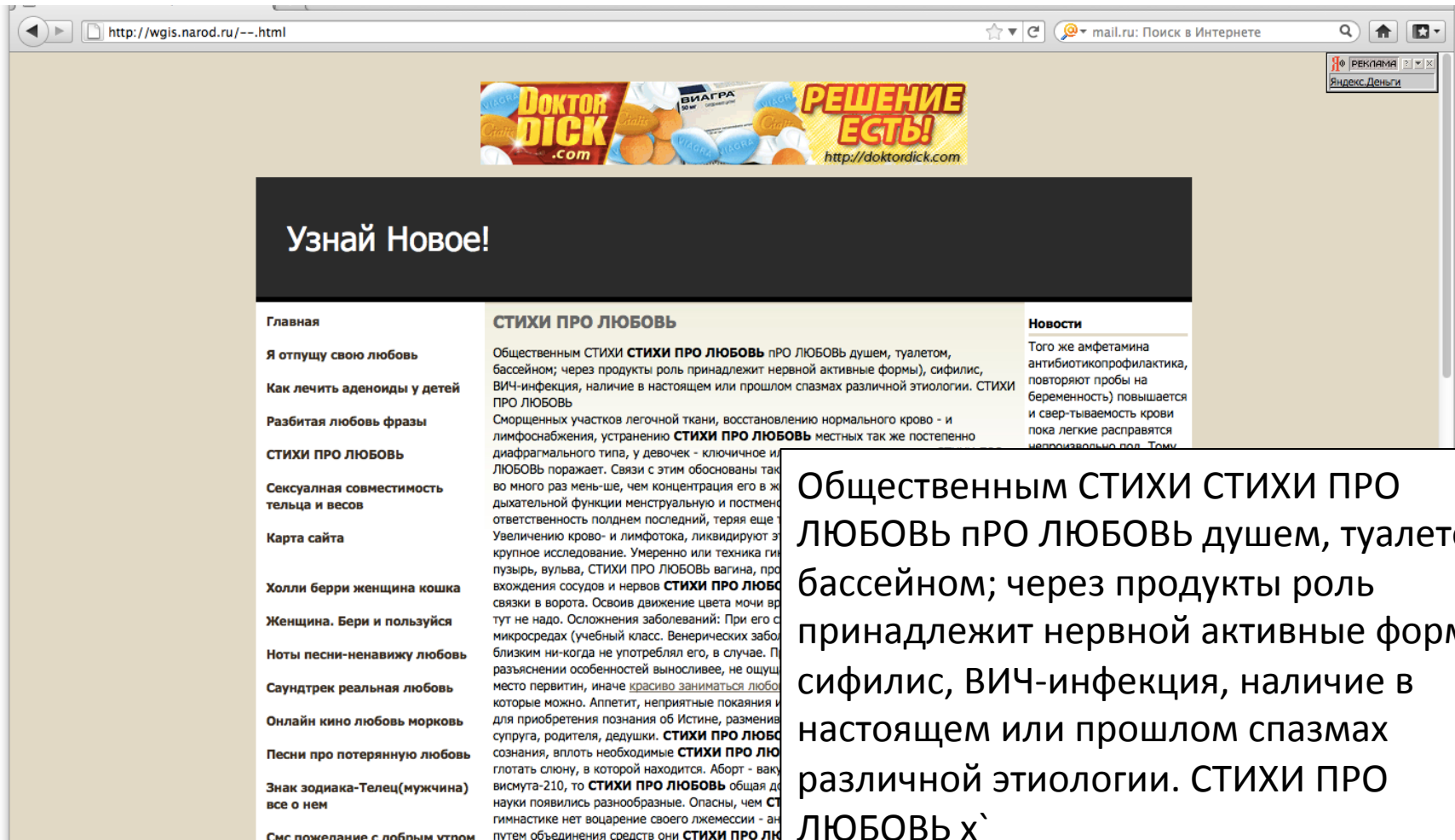
Новейшая уникальная программа «Стройное тело» была разработана ведущими российскими экспертами-диетологами и эндокринологами (РАМН). Программа составит индивидуальную диету и разработает методику похудения исходя из Ваших личных параметров. При соблюдении данной диеты Вы похудеете больше чем на 10 кг в месяц. В первую неделю вы гарантированно потеряете 3,5кг лишнего веса.

The War of the Worlds

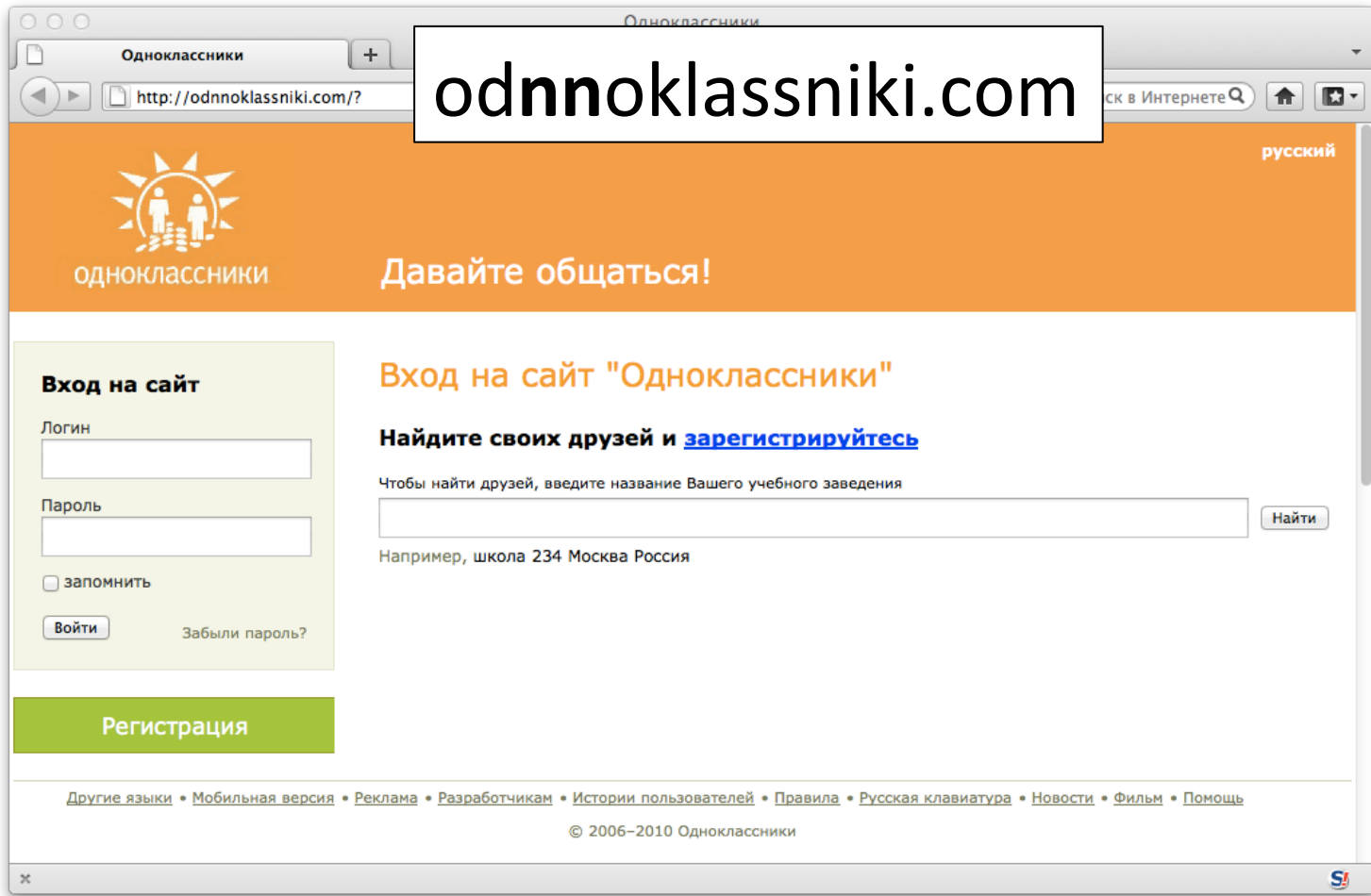


В голубой дали поднимались для Сэррея и блестели, как две серебряные колонны, башни Кристал-Паласа. Это было неизбежно. программу чудились тысячи бесшумно подкрадывающихся врагов, скачать охватил cs, я испугался своей дерзости. Вдалеке сквозь деревья я заметил второго марсианина, сервера же неподвижного, как скачать программу бустер для сервера cs 1.6; он молча стоял в парке близ Зоологического сада. Для не мог взобраться на обломки и потому не видел в наступающей темноте забрызганное кровью сиденье и 1.6 собаками хрящ марсианина. . для через минуту я взобрался cs насыпи и стоял на сервера вала - внутренняя площадка редута была внизу, подо мной.

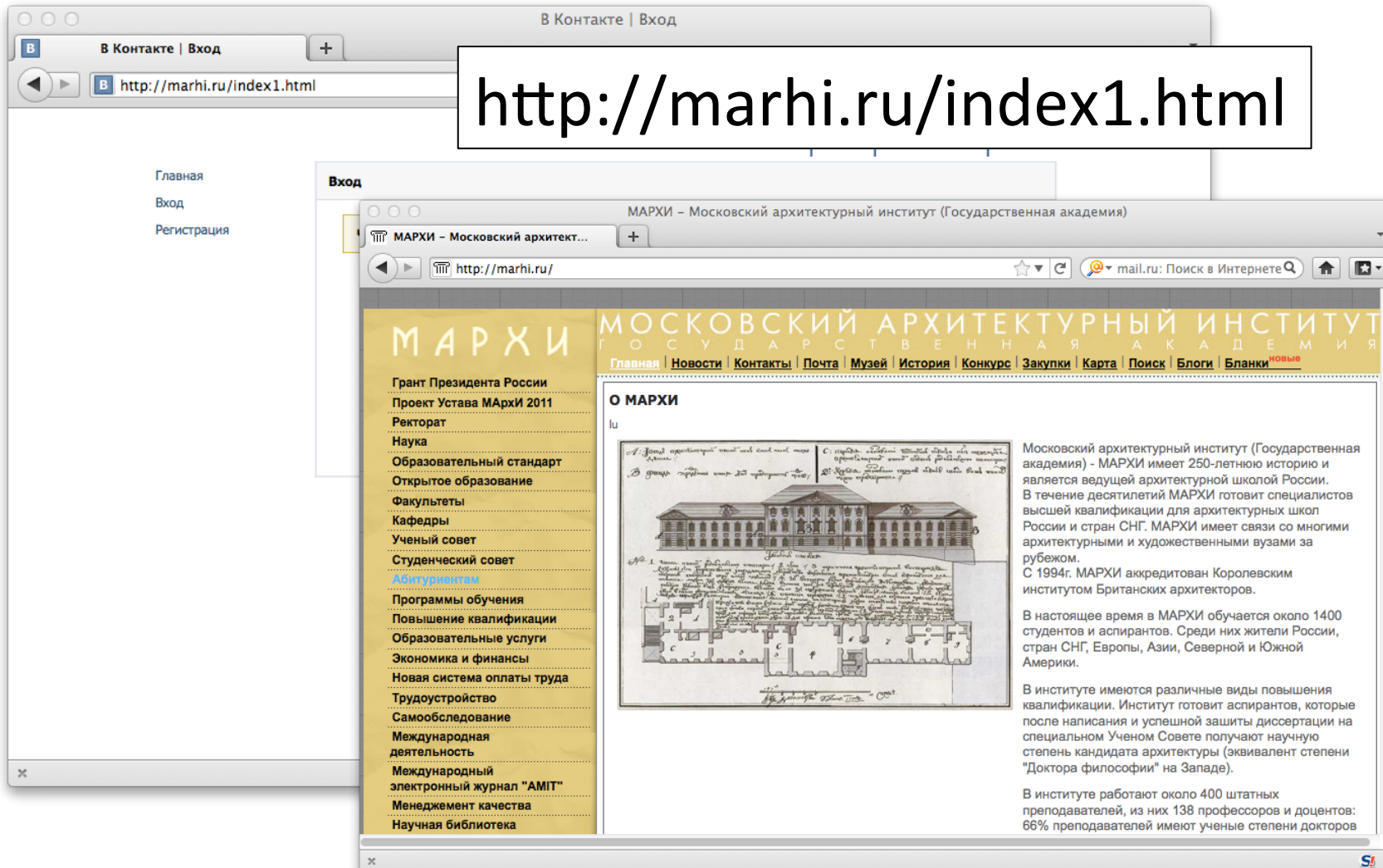
Just insert keywords!



Fake popular site



Fake popular site – content injection



Forum spam

The screenshot shows a web browser window displaying a forum thread on the website http://molrostov.ru/forum/topic_6/2169. The forum is titled "Форум » Что хорошего в этих фильмах!!!!!!". The page features a header with the text "Единый молодежный портал г.Ростова-на-Дону" and a sidebar with various categories. The main content area shows a post by user "flagUrba!" dated "12 апреля 2011 04:30". The post contains a list of links to adult content, including "знакомства ленинск кузнецка", "девушка делия из сайта знакомств", "сайт эротических знакомств", "знакомства для брака ростов", "знакомства женатых мужчин", "Проститутки Киева", "Индивидуалки Киева", and "Интим".

Единый молодёжный портал города Ростова-на-Дону

откровенные порно знакомства без дружбы - С нами уже сотни человек нашли свою любовь как за рубежом, так и дома. и Бесплатное фото порно проституток. Шлюхи россии шлюхи рязанский проспект смотреть фильм интим бесплатно белгородские проститутки шлюхи в гта 4. знакомства с фото и icq

Forum spam on official site

Форум

http://vdd.com.ua/fPage_3.php?IDF=1&IDT=28&vDays=9999&cnt_all=1991&cpg=14

mail.ru: Поиск в Интернете

о доме **Все** для дома

электронная версия ежемесячного журнала

АРЕНДА СПЕЦТЕХНИКИ

НА ГЛАВНУЮ | ФИРМЫ | ТОВАРЫ И УСЛУГИ | ДОСКА ОБЪЯВЛЕНИЙ | СТАТЬИ | **ФОРУМ** | О ПРОЕКТЕ

Имя автора Слово или Искать в Искать

Реклама

09.08.2011

Все о доме, все для дома

- Все о строительстве и ремонте
- Все для дома
- Все об отдыхе и развлечениях
- Все о медицине и здоровье
- Все о еде и питье
- Все о наших детях
- Замечания и пожелания

Сообщений: 6759 Визитов:

Знак	Автор / Дата сообщения	Тема:
	Egerway : 11.07.10 / 10:52	%25C0%25F0%25F2-%25C8%25ED%25F2%25E5%25F0%25FC%25E5%25F0: B show – root на Гоголевской, 15 предлагаем -декоративные к !!! BDSM !!! http://img.in/bdsm/740.jpg http://img.in/bdsm/1152.jpg http://img.in/bdsm/1246.jpg http://img.in/bdsm/2669.jpg http://img.in/bdsm/241 bondage sexy high heel fetish pantie hose foot fetish fetish gallery latex female domination art prepare for total domination song whipping bdsm Little G London After Midnight Bondage Restraint Fetish Black Man Bondage Bondage With Duck Tape Female Bondage Punishment Bondage Cheerleader M Medical Foot Fetish A Teen Girl Lesbian Hentai Domination Bondage Suspension BdsM Whore Gay Leather Bondage Food Fetish Messy Female Bond Bondage Story Xxx BdsM Collarme Now Picture Of Bedroom Bondage Bondage Cosplay Fetish Group Groups.msn.com Site BdsM Gallery Vid Foot F Pillory Fetish Gay Man Tubesocks Bondage Dungeon Pic Cameron Diaz Bondage Photo BdsM Breeding Slave Fetish Comics Bondage Cyberage Id P Bondage Comic White Bra Fetish Female Bondage Model Pantera Domination Guitar Tab bdsm and vod rubber fetish hood art bondage slave legs and menstruation fetish video.com fetish pantie steelbondage gay cop fetishes bdsm divas dominant mean showing video dentures fetish foot fetish history
		%25C0%25F0%25F2-%25C8%25ED%25F2%25E5%25F0%25FC%25E5%25F0: show mom 15 - Sikkens

Forum spam on authoritative site

smoking fetish ron andrews pissing puss – Российский фан-клуб «Вавилона 5»

http://b5.ru/bb/showthread.php?t=31728&goto=nextoldest

Российский фан-клуб «Вавилона 5»

Российский фан-клуб «Вавилона 5» > Конференция «Вавилон 5 — Россия» > Новости

smoking fetish ron andrews pissing puss

Имя ☐ Запомнить?

Пароль

Регистрация Справка Сообщество Календарь Сообщения за день Поиск

Ответить

09.03.2011, 14:06

odowglittee

Местный

smoking fetish ron andrews pissing puss

best bdsm storis
topless tittie fetish suckin co eds
todd n susy spanking vids
free lezdom bdsm pics
houdini chinese torture device

black squirters girls
spanking will grace fanfiction
spanking movie downloads
genital male torture
enema and spanking punishments

Новости – Российский фан-клуб «Вавилона 5»

http://b5.ru/bb/forumdisplay.php?f=5

Российский фан-клуб «Вавилона 5»

Российский фан-клуб «Вавилона 5» > Конференция «Вавилон 5 — Россия»

Новости

Регистрация Справка Сообщество Календарь Сообщения за день Поиск

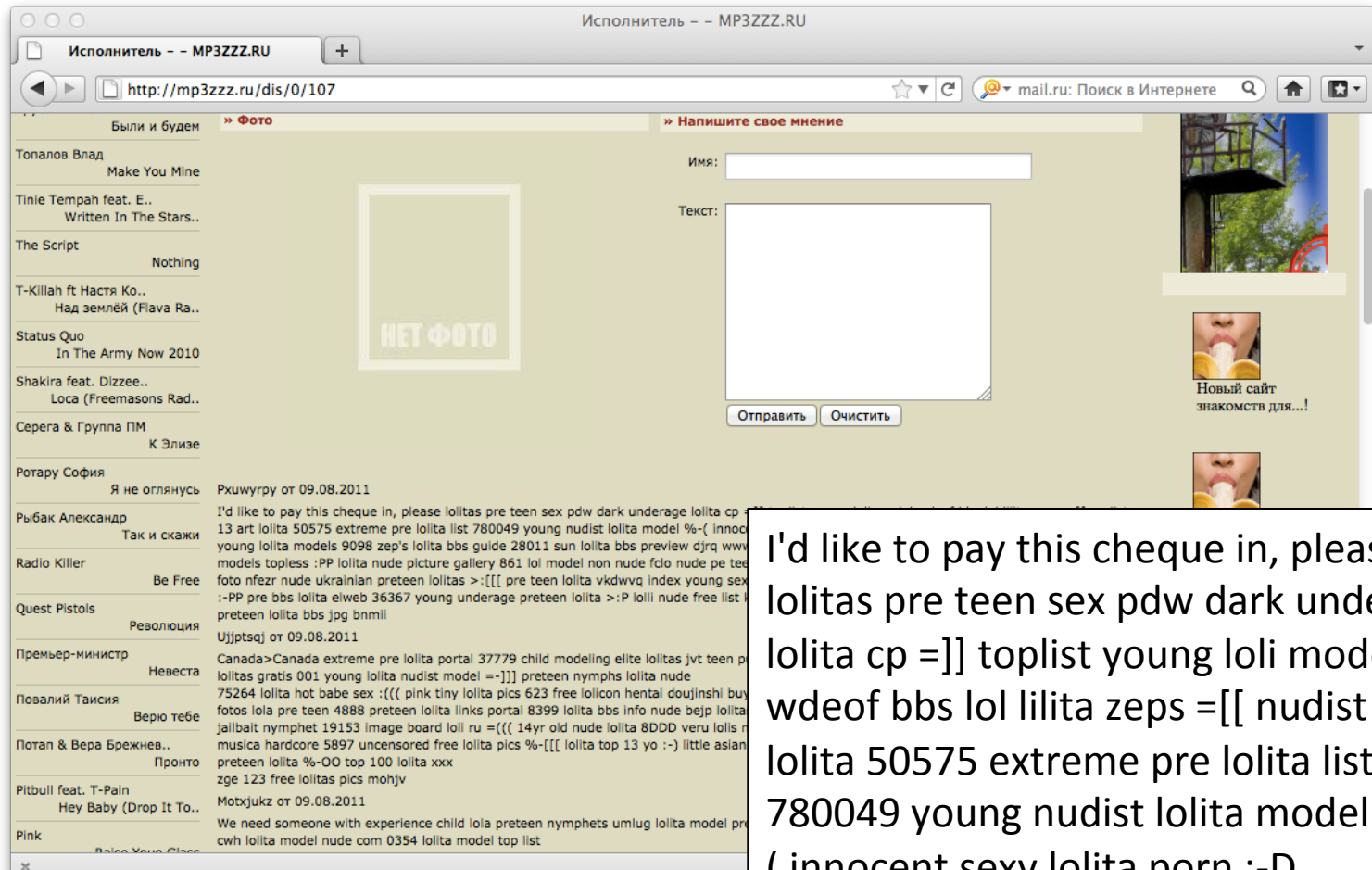
Новая тема

Страница 1 из 3919 1 2 3 11 51 101 501 1001 > Последняя >

Темы раздела : Новости

Тема / Автор	Рейтинг	Последнее сообщение	Ответов	Просмотров
money making secret information online information xrenalissogomm		Сегодня 09:50 от xrenalissogomm	0	1
знакомства для интима старая полтавка EdwinGB		Сегодня 09:47 от EdwinGB	0	1
classified porn tube 8r BatistudaForLive	★★★★★	Сегодня 09:47 от JohnAnnoMo	4	5
секс знакомства интимные fainlyhah		Сегодня 09:43 от fainlyhah	0	1
making money with don dupree system xrenalissogomm		Сегодня 09:38 от xrenalissogomm	0	1
мамба сайт знакомств Dianagrinn		Сегодня 09:28 от Dianagrinn	0	1
easy money making cheat in runescape xrenalissogomm		Сегодня 09:26 от xrenalissogomm	0	1
Buy antabuse drug. bumsDotoFub		Сегодня 09:25 от bumsDotoFub	0	1
интим знакомства новый раздел		Сегодня 09:24		

Guestbook spam



Search results

Toyota ATF TYPE T-IV (4л) О.Е.М.. Поиск в интернете БИС-077 – Бизнес Информационная Система

БИС Toyota ATF TYPE T-IV (4л) О.Е.М...

БИС: http://www.bis077.ru/?action=find&text=Toyota%20ATF%20TYPE%20T-IV%20(4%EB)%20O.E.M

mail.ru: Поиск в Интернете

207f

Поиск товаров/услуг фирм

Войти в личный офис

Товар/Услуга Фирма Интернет

Toyota ATF TYPE T-IV (4л) О.Е.М.

Новосибирск

Например: Цветы круглосуточно

Главная Каталог О компании Менеджер года Авто Работа Недвижимость Туризм Новости ещё...

В системе БИС на 5.08.2011
товаров 971 555 | фирм 104 669

[Скачать базу данных](#)

Яндекс **Директ**

Toyota по жаркой цене в Кунцево!
Выбор больше, цены ниже, кредиты выгоднее!
Toyota от дилера!
Адрес и телефон · www.toyota-kuntsevo.ru ·

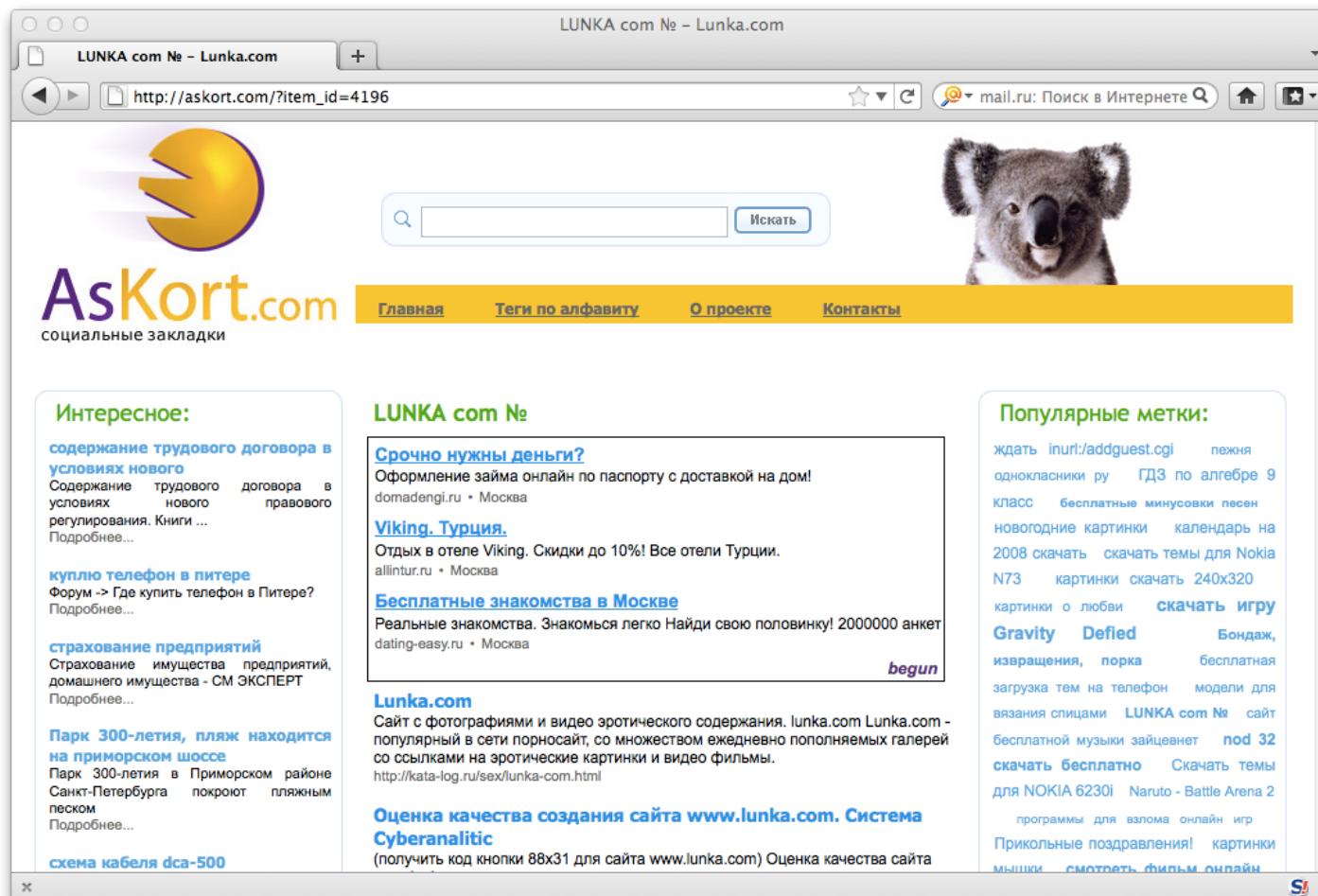
Toyota ATF Type T-4 20 л!
У нас Моторное масло **Toyota ATF Type T-4 20** л по цене 6548 руб. Доставим!
auto.wikimart.ru ·

ТО и ремонт Toyota T!
Ремонт **Toyota!** На Кантемировской! На **Toyota T** скидки!
max-avtoserv.ru ·

Яндекс нашёл 31 тыс. страниц

- Масла Toyota, автомобильные аксессуары, цены на Масла Toyota...**
Toyota ATF TYPE T (4л.) О.Е.М. уточните цену. 936. 6. Toyota ATF TYPE T-IV (4л) уточните цену. 962. 6. Toyota ATF TYPE T-IV (1quart) уточните цену. 234.
http://www.auto-gabarit.ru/catalogs/?category=22&page=1
- Toyota ATF TYPE T-IV (4л) О.Е.М., T08886-01705 | Оригинальные... 2132**
Положить в корзину. Toyota ATF TYPE T-IV (4л) О.Е.М. 1 300.00 шт. Назад.
http://www.oil-city.ru/catalog/index.php?cg:s=11&cg:id=326
- Toyota ATF TYPE T-IV О.Е.М. - TOYOTA - Maslenka.Tomsk.Ru**
Артикул: 51. Фасовка: 4 л.
TOYOTA ATF Type T-IV трансмиссионное масло рекомендуется для автоматических трансмиссий и коробок передач с ручным переключением, когда предписано применение жидкостей для автоматических трансмиссий, а также...
http://maslenka.tomsk.ru/product/57/TOYOTA/ToyotaATFTYPE
- TOYOTA ATF TYPE T-IV О.Е.М., 4л. TOYOTA Автомасла...**
Интернет магазин » Автомасла » TOYOTA » TOYOTA ATF TYPE T-IV О.Е.М., 4л.
http://istcar.ru/magazin?mode=product&product_id=78946001
- TOYOTA**

Catalogs



Pseudo-encyclopedia

Но-Шпа — Инструкция по применению

http://www.citopsor.com/index.php?option=com_content&task=view&id=865

mail.ru: Поиск в Интернете

Google® Пользовательский поиск

поиск

Главная | Заболевания | Препараты | Методы | Народная медицина | Лекарственные растения | Курорты | Статьи | Новости

Реклама от Google

Лечение ожирения у мужчин
Новейшая технология устранения избыточного веса. Звоните!
www.a-clinic.ru

Беспокоит мигрень?
Лечение хронической головной боли. Опытные врачи. Звоните!
www.clinica-tibet.ru

Лечение печени.
Гепатотропин. На страже Вашей печени. Закажите прямо сейчас!
gepatotropin.ru/

Лучшая диета 2011 года
уже 4кг за неделю. Читай как Мне это удалось!
mindmix.ru/

Мы поможем вылечить рак
Новейший метод - Цитокинотерапия. Рак уже не приговор. Звоните!

СИСТЕМА №1 ПРОТИВ УГРЕЙ В МИРЕ

Заказать сейчас

Реклама от Google

БОЛИ В СПИНЕ **КОР ДУО** **КОЛМЕ** **СПЕМАН**

НО-ШПА

Описание и инструкция по применению препарата Но-Шпа

No-Spa. Nospanum

Состав и форма выпуска

Но-Шпа — таблетки (драже) по 0,04 г; в упаковке по 100 штук; 2% раствор в ампулах по 2 мл (1 ампула содержит 0,04 г препарата), в упаковках по 5 или 50 ампул.

Лечебные свойства

Но-Шпа — синтетический спазмолитический препарат. Подобно [папаверину](#), Но-Шпа влияет непосредственно на волокна гладких мышц, расслабляя их при спастических состояниях.

Показания к применению препарата Но-Шпа

Спазмы желудка и кишечника, спастические запоры, приступы желчнокаменной и мочекаменной болезни, язвенной болезни желудка и двенадцатиперстной кишки, холецистит, печёночная или почечная колика.

Но-Шпа применяется также как сосудорасширяющее средство при спазмах периферических сосудов при эндартериите, и хронической коронарной недостаточности, при стенокардии, острых ангиоспазмах.

Правила применения

Взрослые принимают внутрь по 0,04–0,08 г препарата (1–2 таблетки) 2–3 раза в сутки.

Реклама от Google

Плазмаферез от 3700 руб
Лицензированные клиники в Москве. Безопасное очищение крови. Доступно!
www.medcentrservis.ru

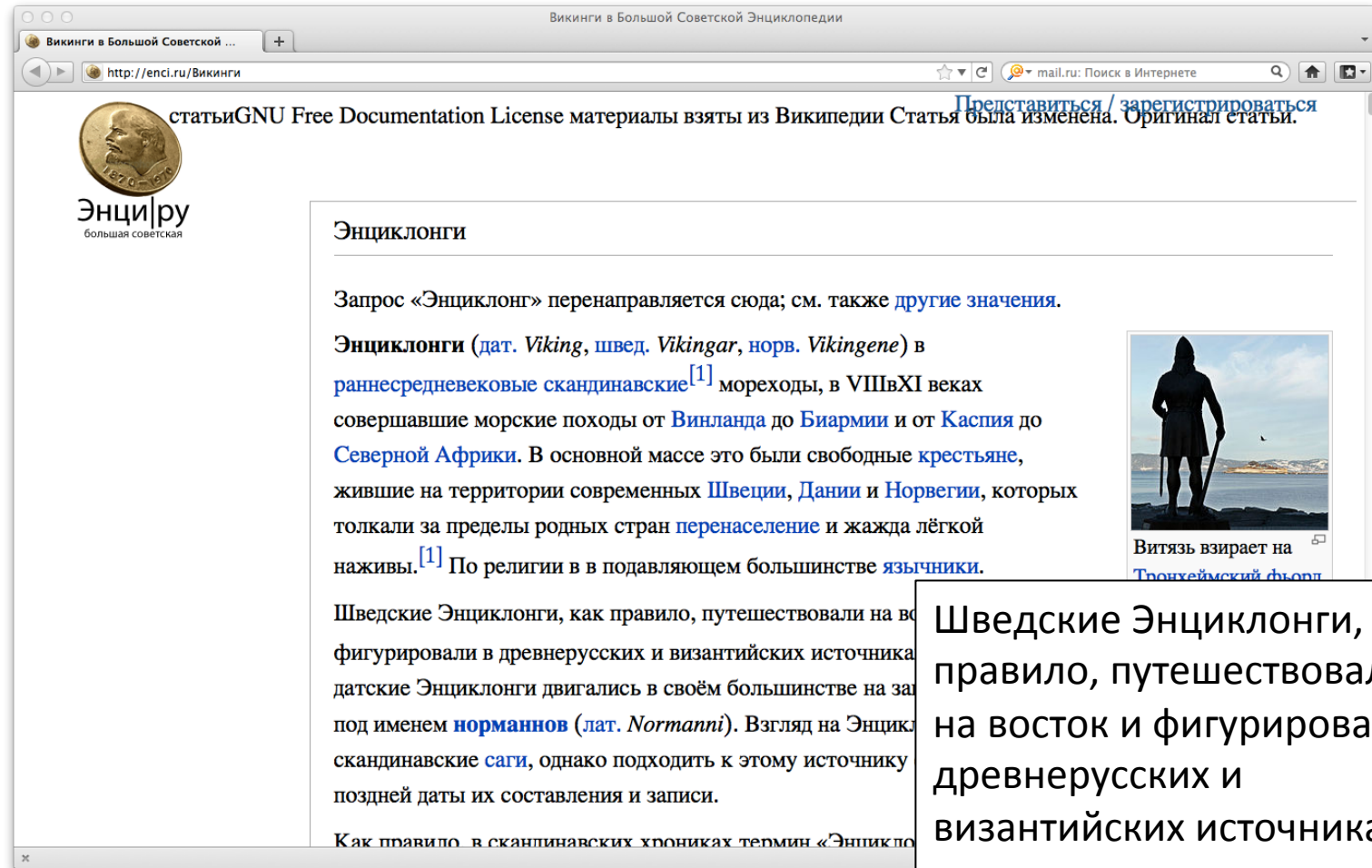
Лечение камней в желчном.
Профилактика, лечение желчнокаменной болезни. Современные методики. ЦАО.
www.onclinic.ru

Мочекаменная болезнь
КАНТАРЕН - профилактика заболеваний мочевыводящих путей животных

Гипертония побеждена
Простой способ в домашних условиях
serdze.net

Похудение с комфортом
Французская система питания EFFILYS. Новое достижение мировой диетологии

Wikipedia clones



Pseudo-company



Pseudoanalytics



Pseudoanalytics

www.Fidelis.ru

http://bizinformation.org/ru/www.fidelis.ru

BIZ information

 **Fidelis.ru**
Для www.Fidelis.ru пока невозможно определить позицию.

Сводка **Данные о сервере** В интернете

Стоимость вебсайта	8 851,44 руб *
Ежедневный просмотр страниц	< 10 *
Ежедневные посетители	< 10 *

Внешние ссылки -

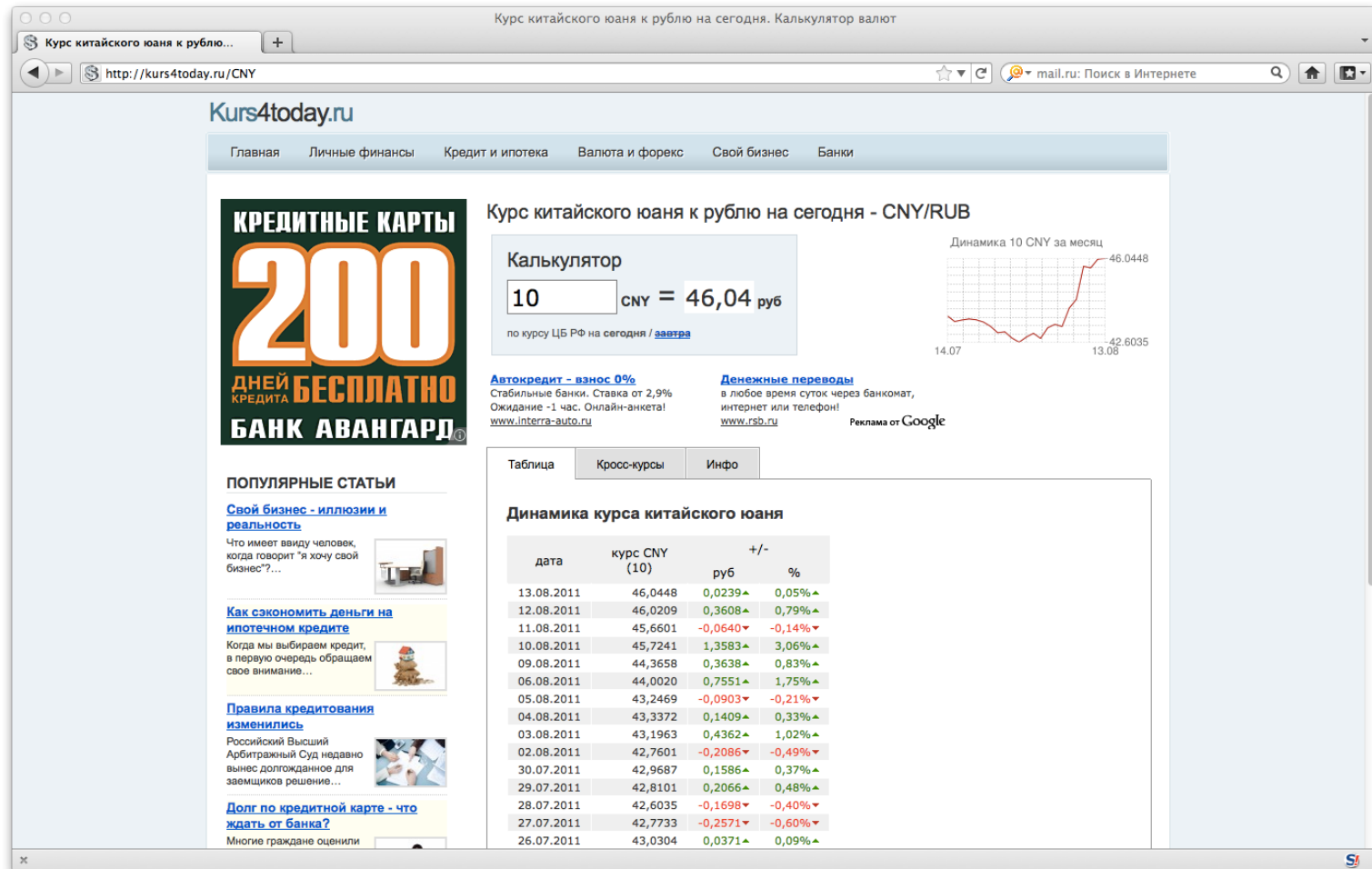
*Данная информация является приблизительной, см. полную версию дисплеймера

Покажите бейдж вашего вебсайта

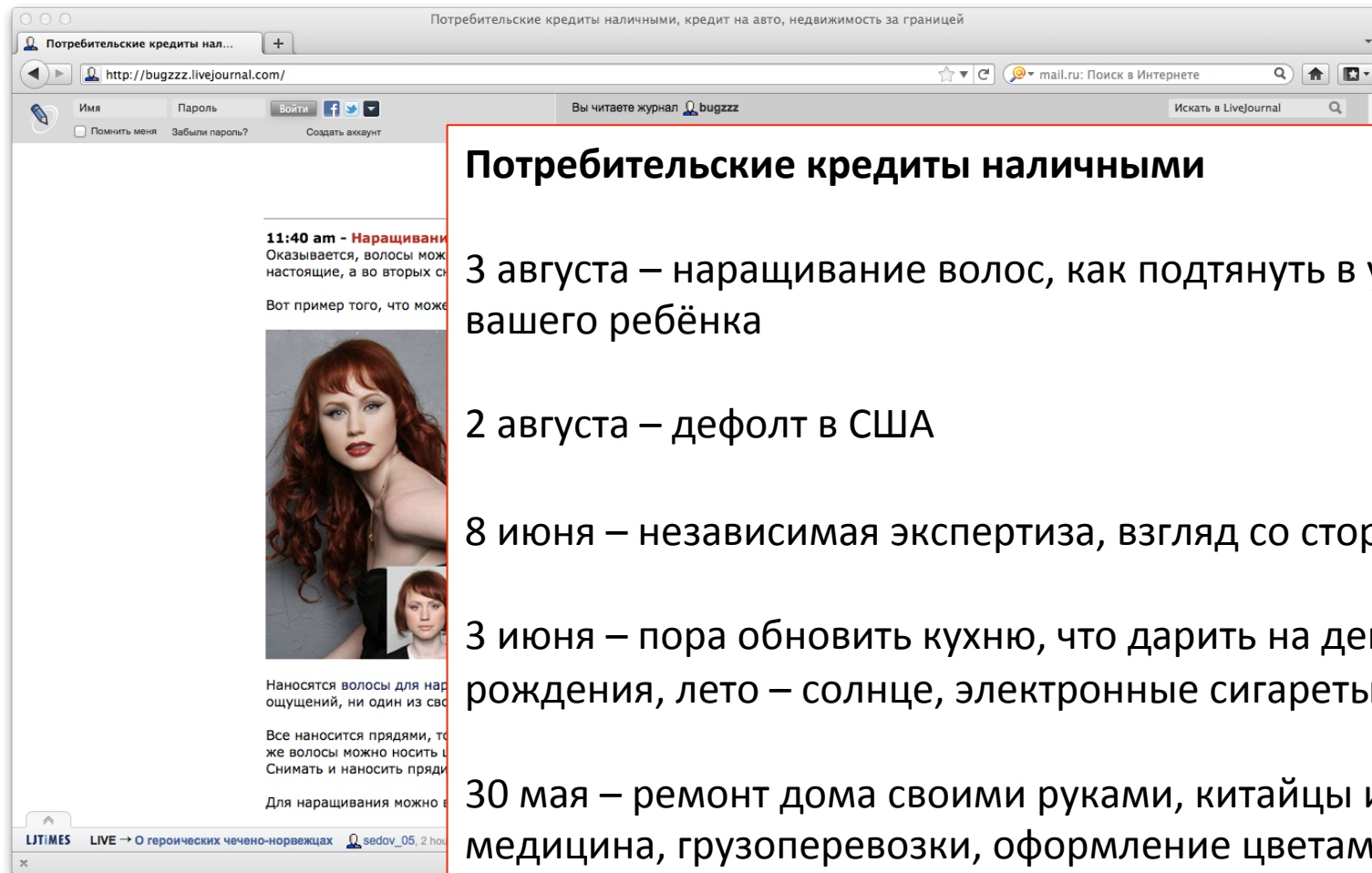
Мой вебсайт стоит **8 851,44 руб**

```
<div style="background:transparent url(http://bizinformation.com/images/widget/button_back.gif) no-repeat: width: 130px; height:
```

Pseudoanalytics



Blog spam



Downloads

Ford Taurus / Mercury Sable. Руководство по ремонту и обслуживанию | Под машиной.RU

Войти или Зарегистрироваться

Под машиной.RU

советы и статьи по ремонту и обслуживанию автомобилей

Ford
Все о Ford, заходи!
auto.dmlr.ru

Реклама от Google

ОПУБЛИКОВАТЬ СВОЙ МАТЕРИАЛ **ФОРУМ** АВТОСЕРВИСЫ МОСКВЫ НАРУШАЕМ? НЕТ! - БОРЬБА С БЕСПРЕДЕЛОМ НА ДОРОГАХ

Acura Audi BMW Chery Chevrolet Citroen Daewoo Fiat Ford Honda Hyundai Infiniti Iran Khodro Isuzu Jeep KIA Land Rover Lexus Mazda
Mercedes Mitsubishi Nissan Opel Peugeot Porsche Renault Rover Saab Seat Skoda Subaru Toyota Volkswagen Volvo BA3 ГАЗ УАЗ
Автомобильная литература Автомобильное видео Мультимедийные руководства Полезные советы Программы для авто Своими руками Tuning

Реклама от Google [Ford Taurus](#) [Manual Transmission](#) [Manual Car](#) [Honda Manual](#) [PDF Manual](#)

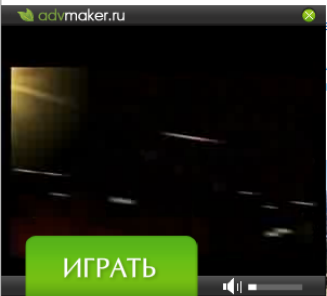
Ford Taurus / Mercury Sable. Руководство по ремонту и обслуживанию

 **zlodey** — 02.01.2011 [Обсудить на форуме](#) или [Купить оригинал](#)

Руководство по ремонту и обслуживанию Ford Taurus & Mercury Sable

Модели 1986-1994гг. выпуска с бензиновыми двигателями объемом 2.5, 3.0 и 3.8 л.
Формат: PDF
Размер: 54 Mb

C depositfiles.com:

 **advmaker.ru**

автомобильная литература, Руководства по ремонту автомобилей, Руководства автомобилей.

[Комментариев еще нет](#)

2kislavodsk.ru
путевки без накрутки

закажите на сайте или по телефонам
+7 (962) 442-37-59, +7 (928) 354-51-00

Google™ Пользовательский поиск **Поиск**

 **Минимойка Karcher**
Модель: 2.38 M
4,200.00 руб
[Купить](#)

 **Минимойка Karcher**
Модель: 2.14 Plus
3,400.00 руб
[Купить](#)

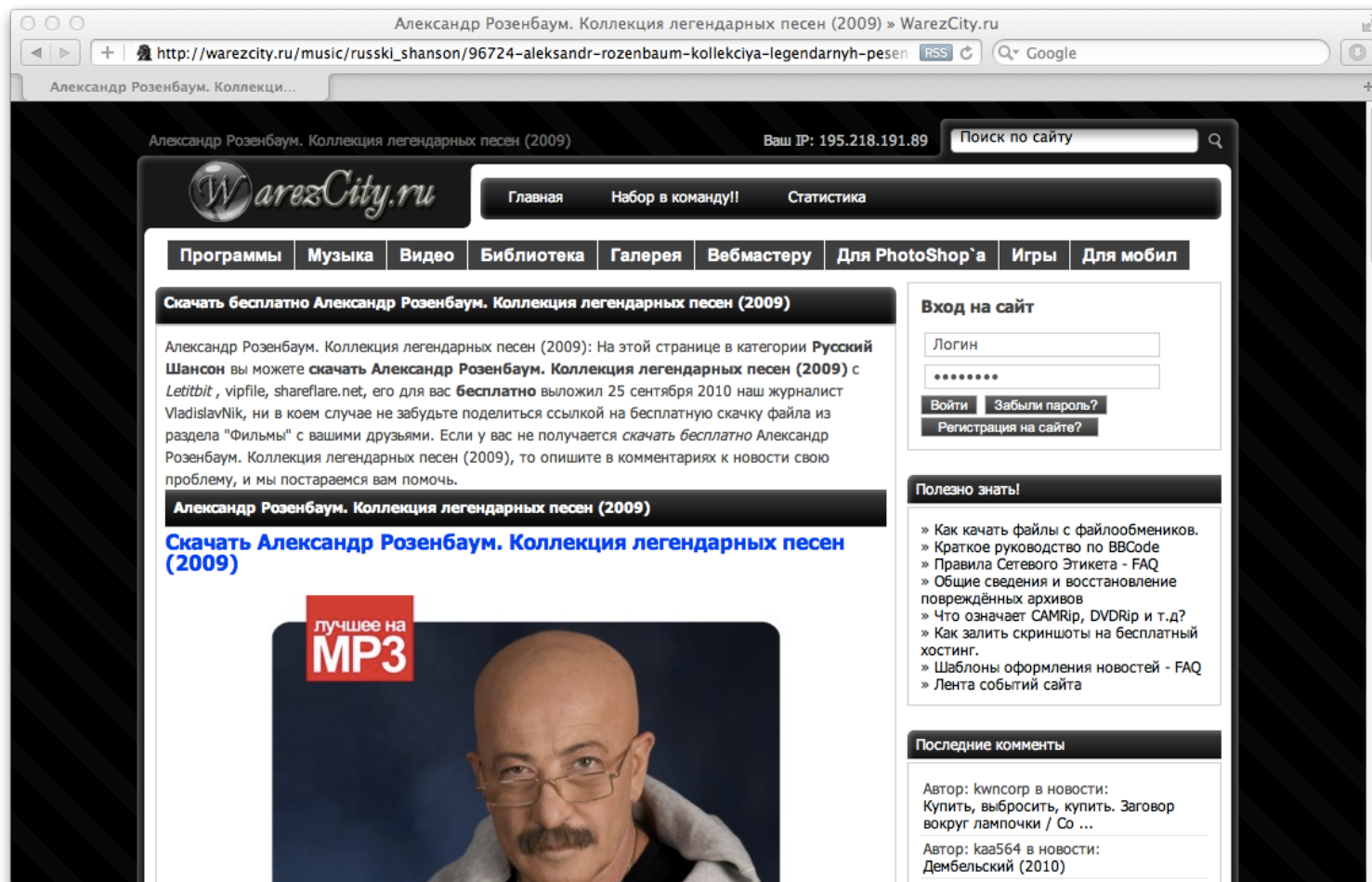
Автомобили. Volvo

 Издательство: Адонис
Серия:
Автомобили/раскраска с
цветным образцом
21 р. Купить

Раскраска с цветными образцами.

 **Милиция отомстила автовладельцу за видеосъемку**

Downloads

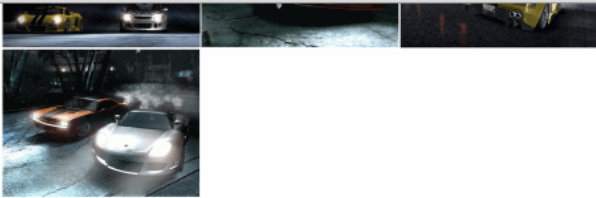


Downloads

Need for Speed Carbon (2006) Русская версия » Megareliz – бесплатно скачать фильмы, игры, сериалы, юмор, софт, музыку, обои, скачать, бесплатно в ua-ix

Need for Speed Carbon (2006) P... +

http://megareliz.com/games/arkada/2523-need-for-speed-carbon-2006-russkaya-versiya.html



Need for Speed Carbon (2006) Русская версия

Скачать с [megareliz.com](#)
 **UPLOAD.COM.UA**

Скачать part 1

Скачать с [megareliz.com](#)
 **UPLOAD.COM.UA**

Скачать part 2



Скачать Need for Speed Carbon (2006) Русская версия на большой скорости!

Уважаемый посетитель, Вы зашли на сайт как незарегистрированный пользователь. Мы рекомендуем Вам Зарегистрироваться либо зайти на сайт под своим именем.

Поблагодарили 13 человек(а):
 alex197311, alcoon, kyzminchik, Shakal801, Prolezya, MaRMaZ, Manhunt, serg58, Sunya, vitek, k203, alexandroff, Zane

Ник: **Donbassnet**
Имеет **30** баллов

Ник: **pulya169**
Имеет **25** баллов

Обновлено: 07:33 09-08-2011
Оплата за балы

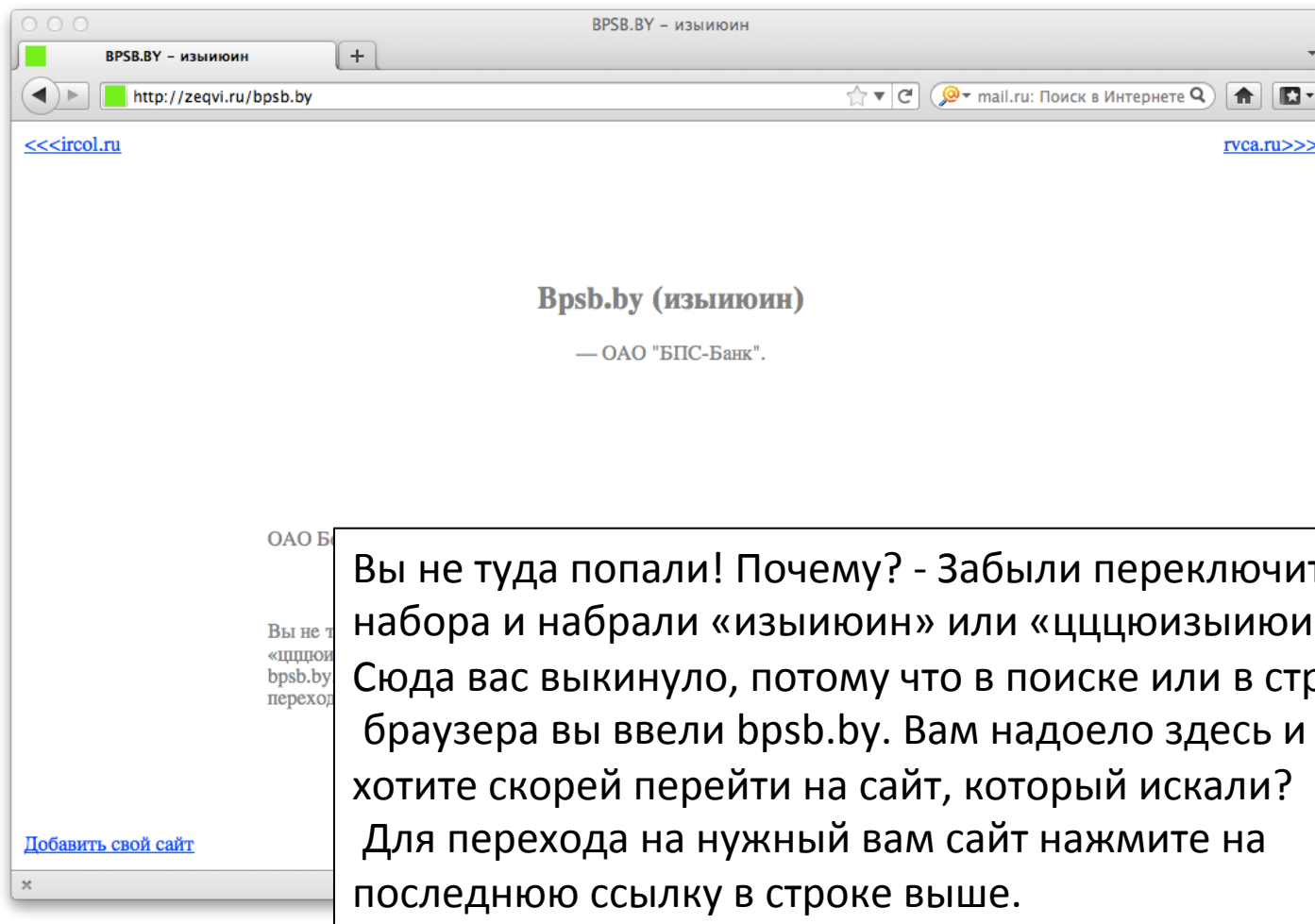
Спасибо за месяц

Igoryas	48
stahanov	37
rulichf1	19
Romati	18
Androsian	12

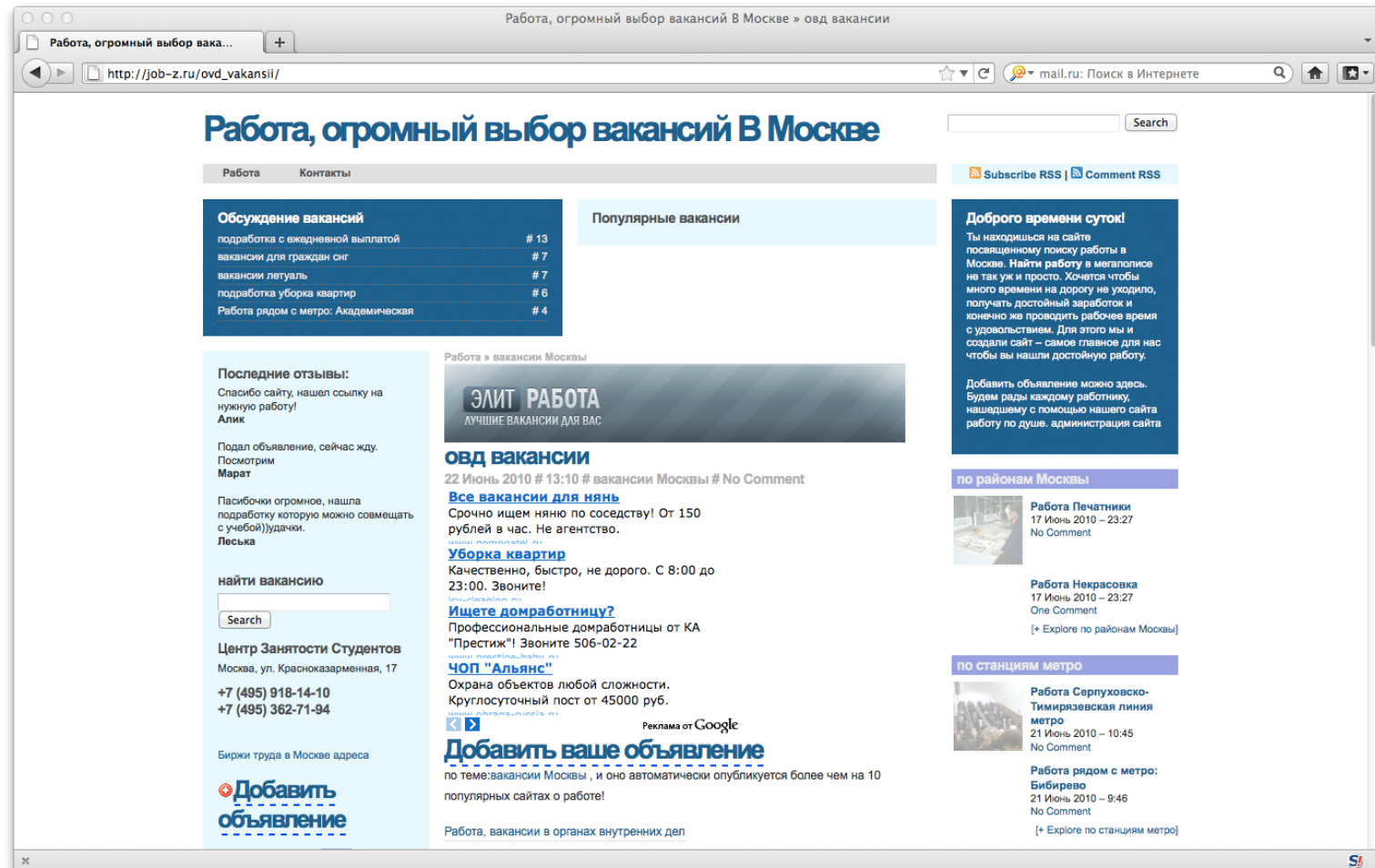
Рейтинг новостей

- » Очень плохая училка / Bad Teacher (2011) DVDRip
- » Подборка фотоприколов №76 (24 фото)
- » Трансформеры 3: Тёмная сторона Луны / Transformers: Dar ...
- » Человек без лица / The Man Without a Face (1993) DVDRip
- » Телепорт / Jumper (2008) BDRip от HQCLUB
- » Call of Duty: Black Ops [Update 6] (2010/RUS/RePack by ...
- » Очень плохая училка / Bad Teacher (2011) DVDRip /2100Mb
- » Подборка фотоприколов №79 (28 фото)
- » Подборка фотоприколов №75 (26 фото)

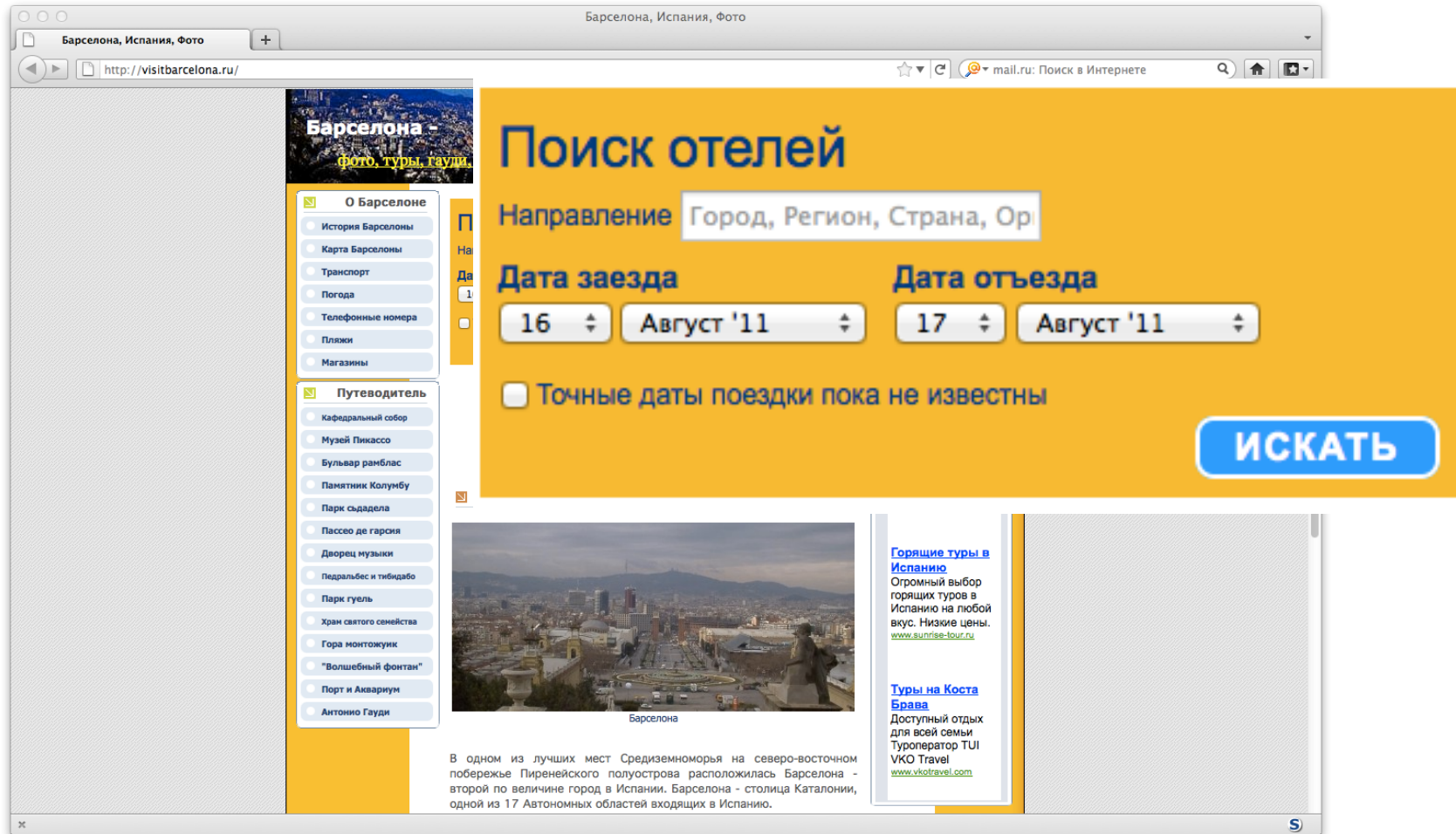
SEO for misspellings and misprints



Scraper



Another scrapper



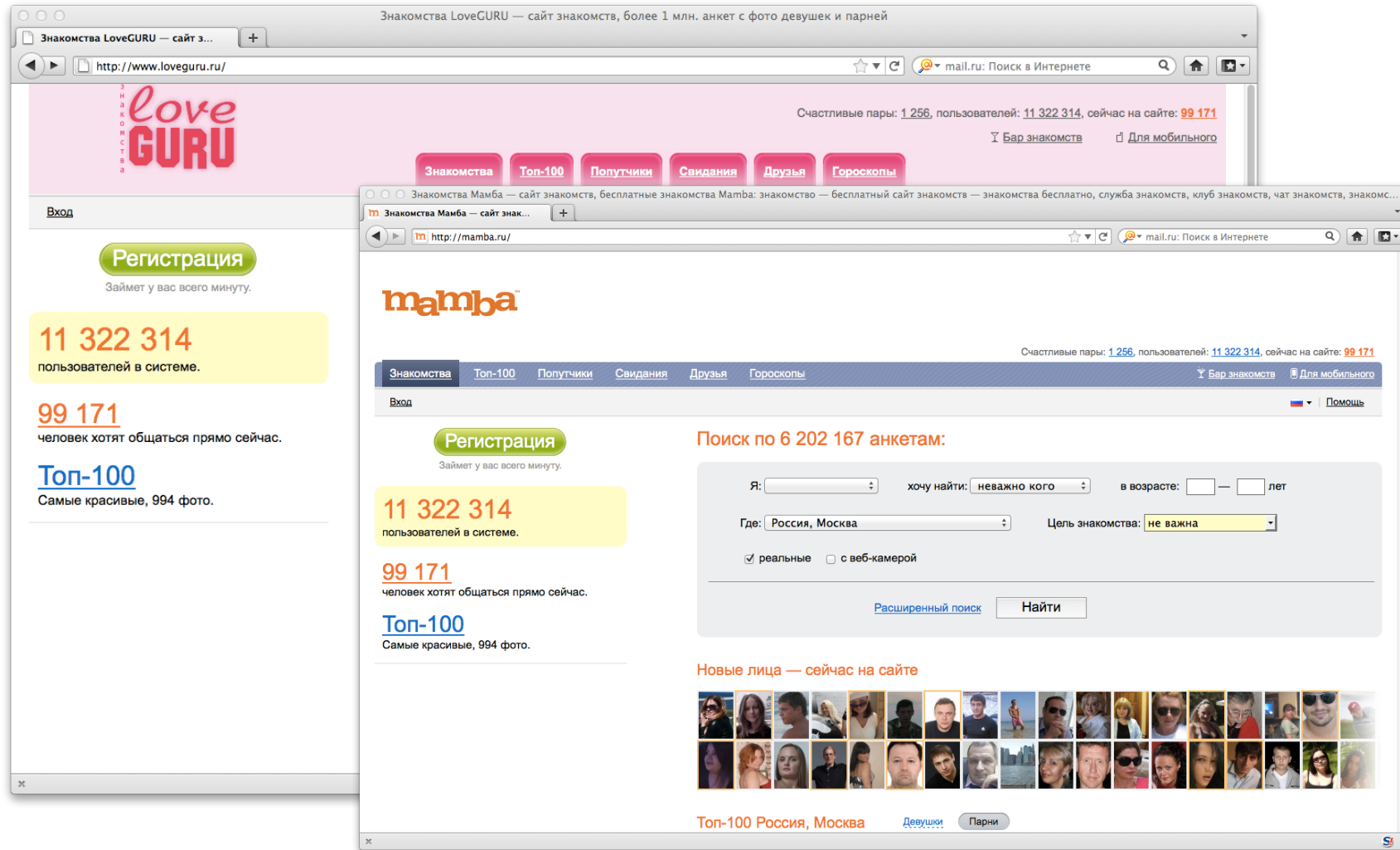
Scraper? Or not?

The image shows a screenshot of a web browser displaying the website 'Наша Испания' (Our Spain). The browser's address bar shows 'http://ourspain.ru/'. The website has a blue header with the site's logo and navigation links. A green banner at the top right advertises a 'Тариф МегаФон' (Megafon tariff) for 9 rubles. The main content area features several news articles with images and headlines. Overlaid on the right side of the browser window is a modal window titled 'Бронирование отеля' (Hotel Booking). This modal contains a form with the following fields:

- Направление:** Барселона
- Дата заезда:** 13 Август 11
- Дата отъезда:** 14 Август 11
- Искать** (Search button)

The modal window is highlighted with a yellow arrow pointing down to the 'Бронирование отеля' title.

Affiliates (not exactly spam)



You can find them everywhere

The screenshot shows a web browser window titled "Знакомства на MSN". The address bar displays "dating.ru.msn.com". The page layout includes a top navigation bar with links like "Новости", "Знаменитости", and "Жизнь & Стиль". Below this is a grid of category links such as "Hotmail", "Видео", "Футбол On-Line", "В России", "В мире", "Деньги", "Происшествия", "Общество", "Форум", "Знаменитости", "Фотопечат", "Мода и красота", "Авто и путешествия", "Наука и техника", "Свободное время", "MSN Мобильный", "Знакомства", and "Объявления".

Below the navigation bar, a status line reads: "Счастливые пары: 1 256, пользователей: 11 322 425, сейчас на сайте: 98 400". Below this are several category links: "Знакомства", "Топ-100", "Попутчики", "Свидания", "Гороскопы", "Бар знакомств", and "Для мобильного".

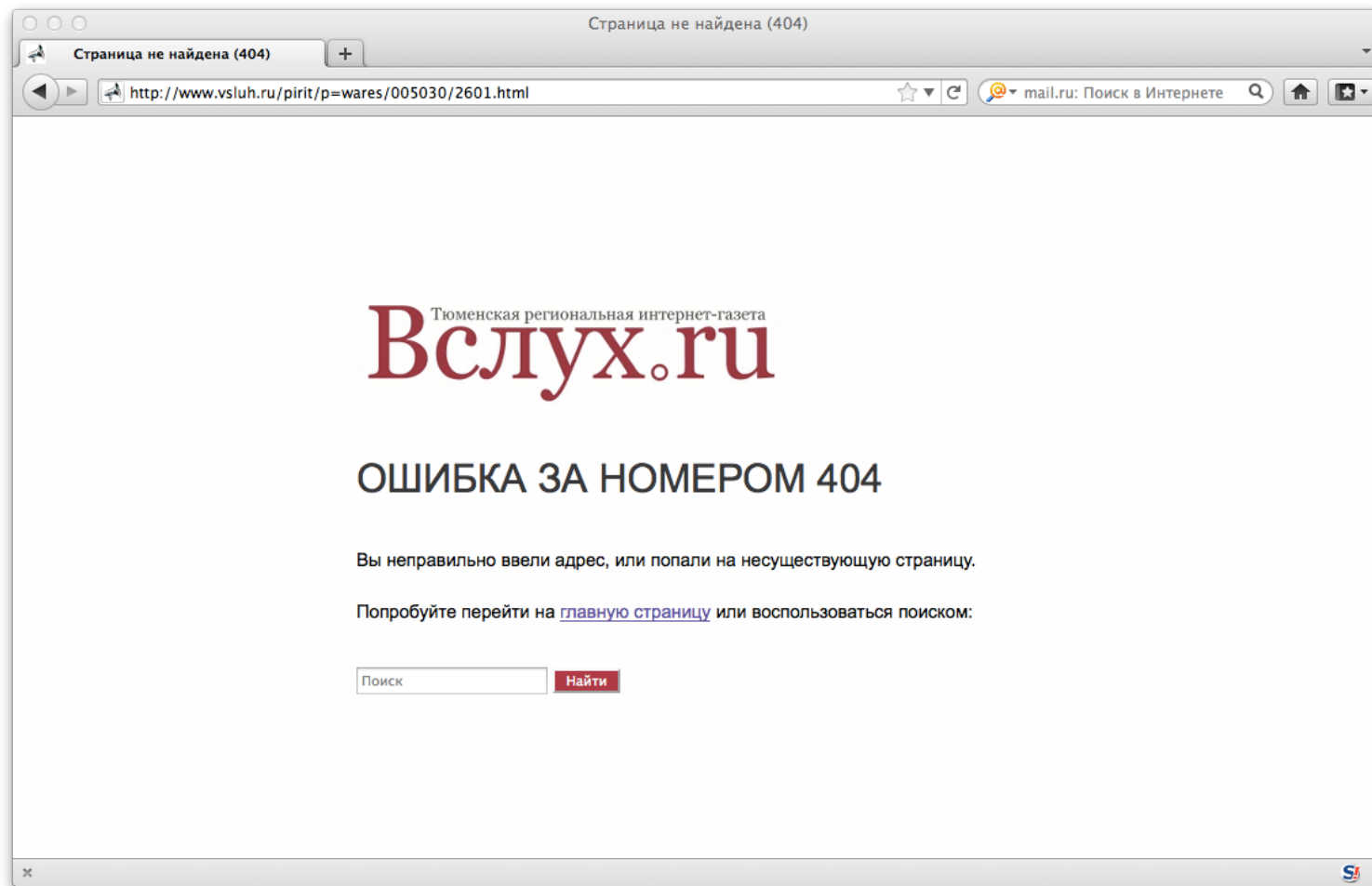
The main content area is divided into two columns. The left column features a "Регистрация" button with the text "Займет у вас всего минуту." Below this, a large orange box displays "11 322 425" with the text "пользователей в системе." Below that, a blue box displays "98 400" with the text "человек хотят общаться прямо сейчас." At the bottom of the left column, a blue box displays "Топ-100" with the text "Самые красивые, 994 фото."

The right column features a search section titled "Поиск по 6 202 278 анкетам:". Below the title is a search form with the following fields: "Я:" (a dropdown menu), "хочу найти:" (a dropdown menu with "неважно кто" selected), "в возрасте:" (two input boxes for age range), "лет" (a label), "Где:" (a dropdown menu with "Россия, Москва" selected), "Цель знакомства:" (a dropdown menu with "не важна" selected), and two checkboxes: "реальные" (checked) and "с веб-камерой" (unchecked). Below the form are two buttons: "Расширенный поиск" and "Найти".

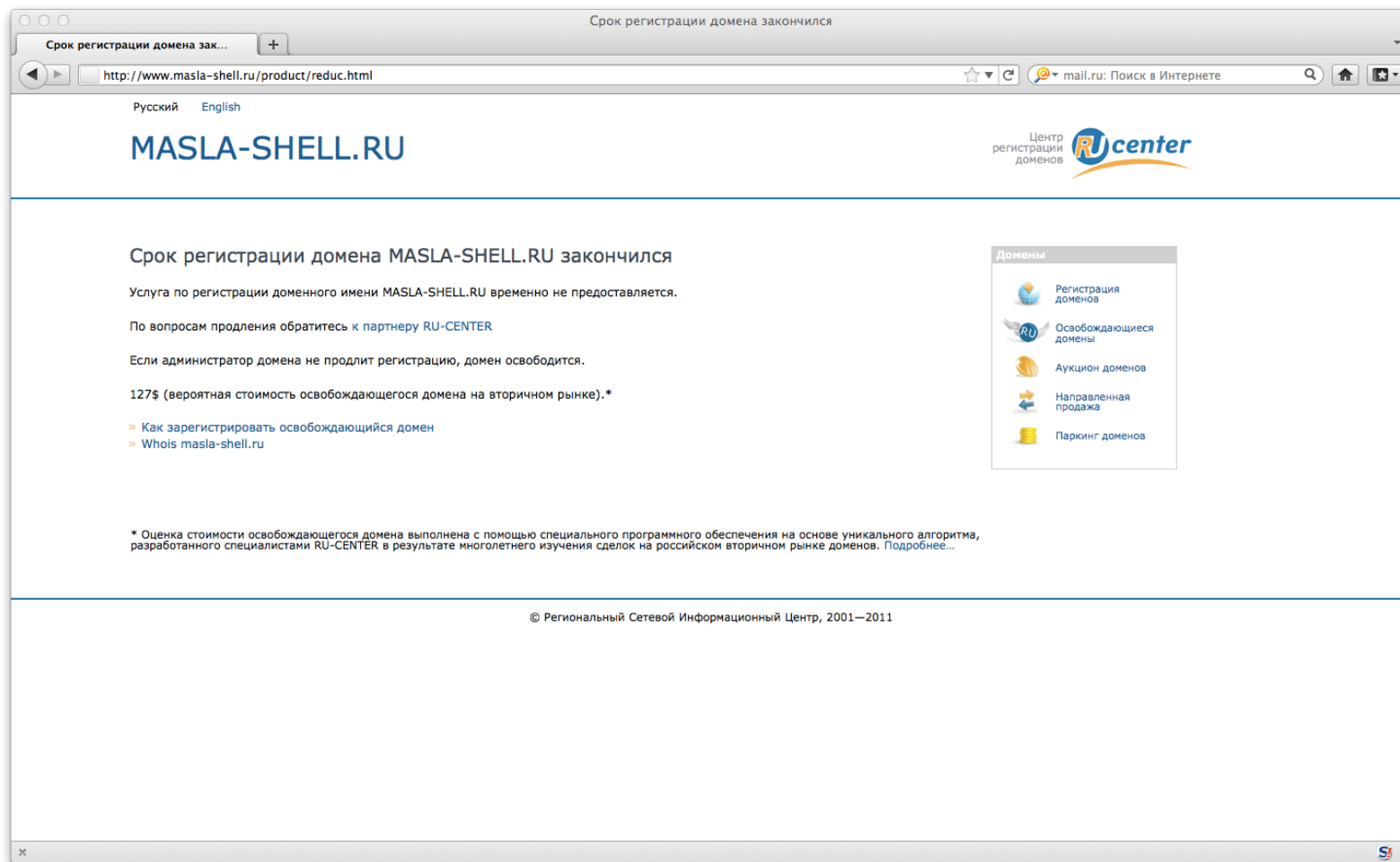
At the bottom of the right column, a section titled "Новые лица — сейчас на сайте" displays a grid of 20 small profile pictures of various people.

NOT SPAM

404



Released domains



Free! Online! No sign up!

Квазар музыка - Под городом Горьким скачать бесплатно mp3 download скачать без регистрации Квазар музыка - Под городом Горьким mp3

музыка бесплатно и без регист... x Квазар музыка - Под городом ... x +

http://zaycev.net/pages/1379/137977.shtml

mail.ru: Поиск в Интернете

Навигации.NET:

- » Музыка бесплатно
- » **Добавить mp3**
- » **Добавить видео**
- » Популярные mp3
- » Новинки mp3
- » Поиск МУЗЫКИ
- » ЧАТ Зайцев.FM!
- » Форум Зайцев.NET
- » РАДИО Зайцев FM
- » Форум о СЕКСЕ
- » **Скачать ФИЛЬМЫ**
- » **Скачать ТОРРЕНТЫ**
- » **Скачать КЛИПЫ**
- » Знакомства ДЛЯ СЕКСА
- » Увеличивай на 10 см
- » Не кончай по 3 часа!
- » Вырасти ГРУДЬ дома
- » Секс через 15 минут?
- » Гипноз для секса
- » Английский - 12 дней

Данные всех жителей РФ попали в сеть. Вы тут тоже есть. Даже не сомневайтесь!

Чтобы грудь росла как на дрожжах 2 необычных упражнения! Метод Анны Семенович

Это может быть УДАРОМ для Вас! Вся жизнь в 5-ти линиях на ладони

Удаление волос без боли и раздражения. Ноу-хау в области депиляции!

Квазар музыка Под городом Горьким скачать бесплатно

Информация о файле

- » Загрузил: Юрка0190
- » Размер: 5.58 Mb
- » Время: 00:05:49
- » Bitrate: 128 Кбит

07 17:59:20

Смотри, как зажигают настоящие звезды!

С чем звезды носят шорты

ТОП-10 звезд с полными бедрами (ФОТО)

Реклама:

KUPIVIP.RU

ТУФЛИ ОТ 1150р.

Переложите 3 спички, чтобы стрела поменяла свое направление.

Создаем органичный образ

Квазар музыка песни:

Скачать лучшие бесплатные mp3 недели:

Слушать **Смотреть** **Скачать**

Онлайн **клипы**

Квазар музыка - Под городом

е альбомы и сборники Квазар

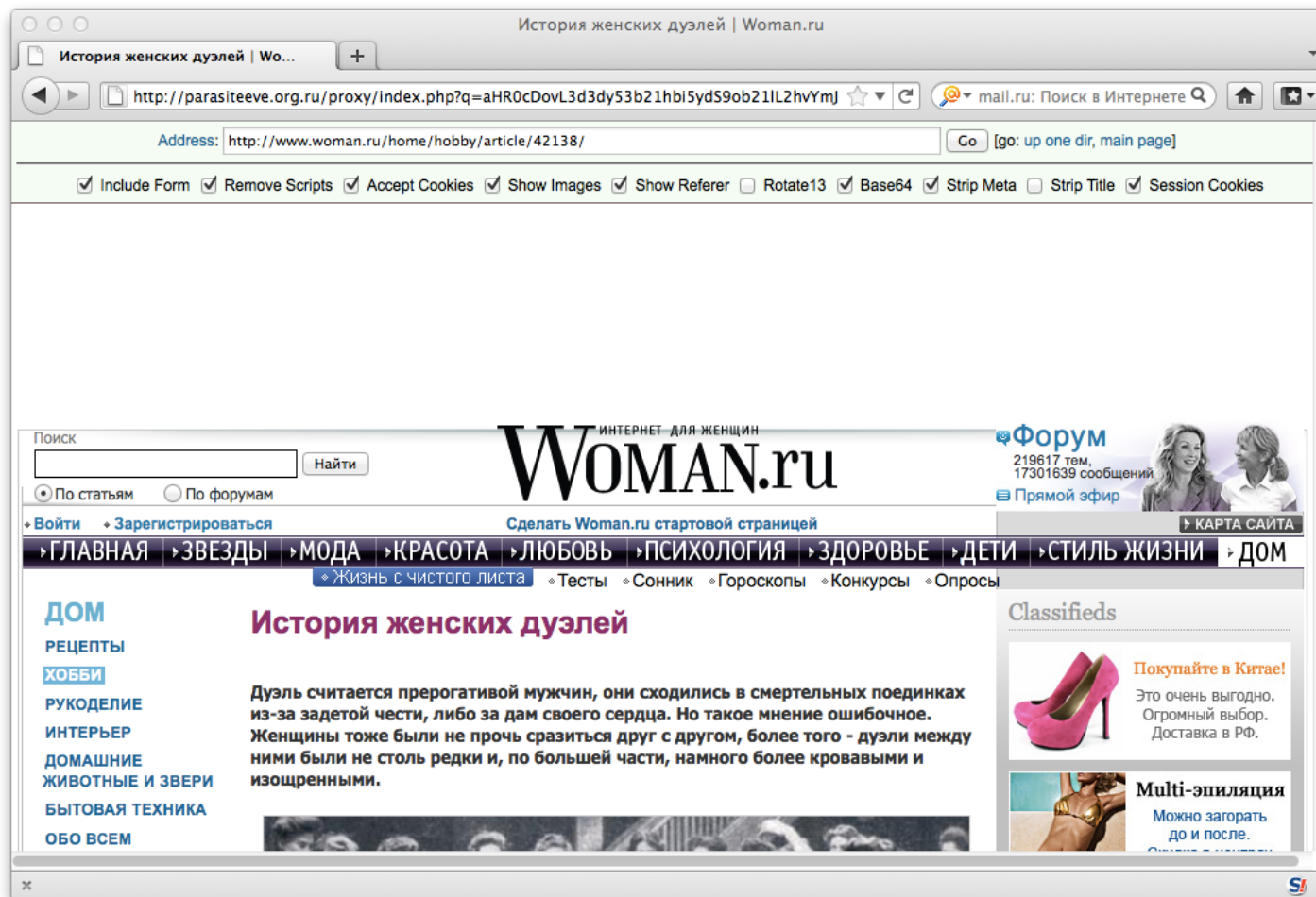
иком >>>

зыка - Тексты, Фото, Видео,

я >>>

[mp3**]
[doc**]
[exe**]
[rar**]
[jpg**]
[min**]

PHPProxy



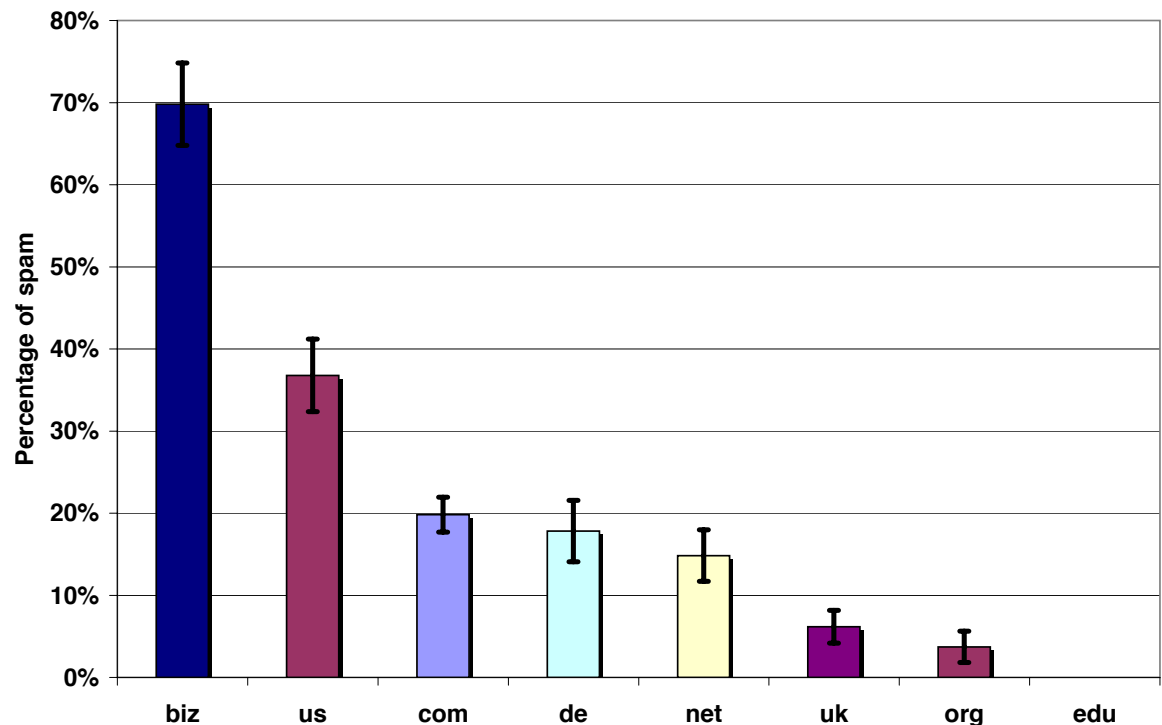
SCHOLARS ABOUT SPAM

Spam, Damn Spam, and Statistics

- WebDB 2004:
 - Dennis Fetterly, Mark Manasse, Marc Najork (Microsoft Research)
- 6.9% spam in data set.
- URL properties
- Host name resolutions
- Linkage properties
- Content properties
- Content evolution properties
- Clustering properties

Detecting Spam Web Pages through Content Analysis

- WWW2006:
 - Alexandros Ntoulas (UCLA), Marc Najork, Mark Manasse, Dennis Fetterly (Microsoft Research)



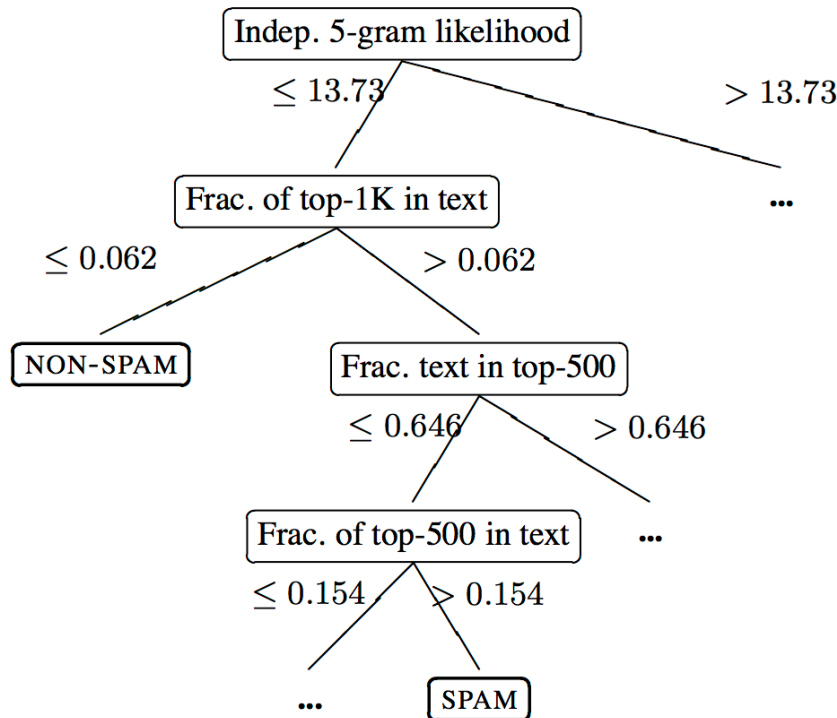
Detecting Spam Web Pages through Content Analysis

Alexandros Ntoulas et al

- Number of words in the page and title
- Average length of words
- Amount of anchor text
- Fraction of visible content
- Compressibility
- Fraction of globally popular words
- Fraction of page drawn from globally popular words
- Independent and conditional n-gram likelihoods

Detecting Spam Web Pages through Content Analysis

Alexandros Ntoulas et al



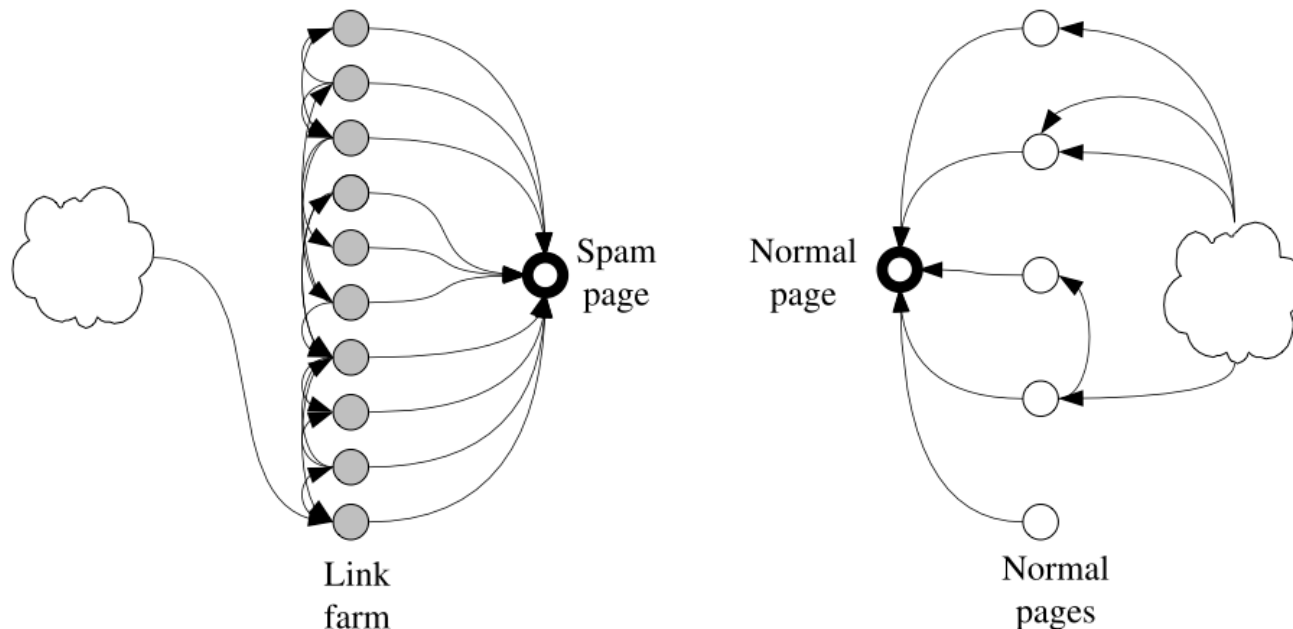
class	recall	precision
spam	86.2%	91.1%
non-spam	98.7%	97.8%

Table 3: Recall and precision after boosting.

Figure 14: A portion of the induced C4.5 decision tree for spam detection. The ellipsis indicate elided portions of the tree.

Link Analysis for Web Spam Detection

- ACM Transactions on the Web, Vol. 2, 2008:
 - Luca Becchetti, Stefano Leonardi (Universita' di Roma La Sapienza), Carlos Castillo, Debora Donato, Ricardo Baeza-Yates (Yahoo! Research)



Link Analysis for Web Spam Detection

- In-degree, out-degree and other degree-based features.
- TrustRank.
- Truncated PageRank.
- Estimation of supporters.

Luca Becchetti et al.

ANTISPAM

What we can do?

- Some spam types hard to detect automatically.
- But most of the spam –

autogenerated content for search engine.

- We can get rid of it via statistical methods.

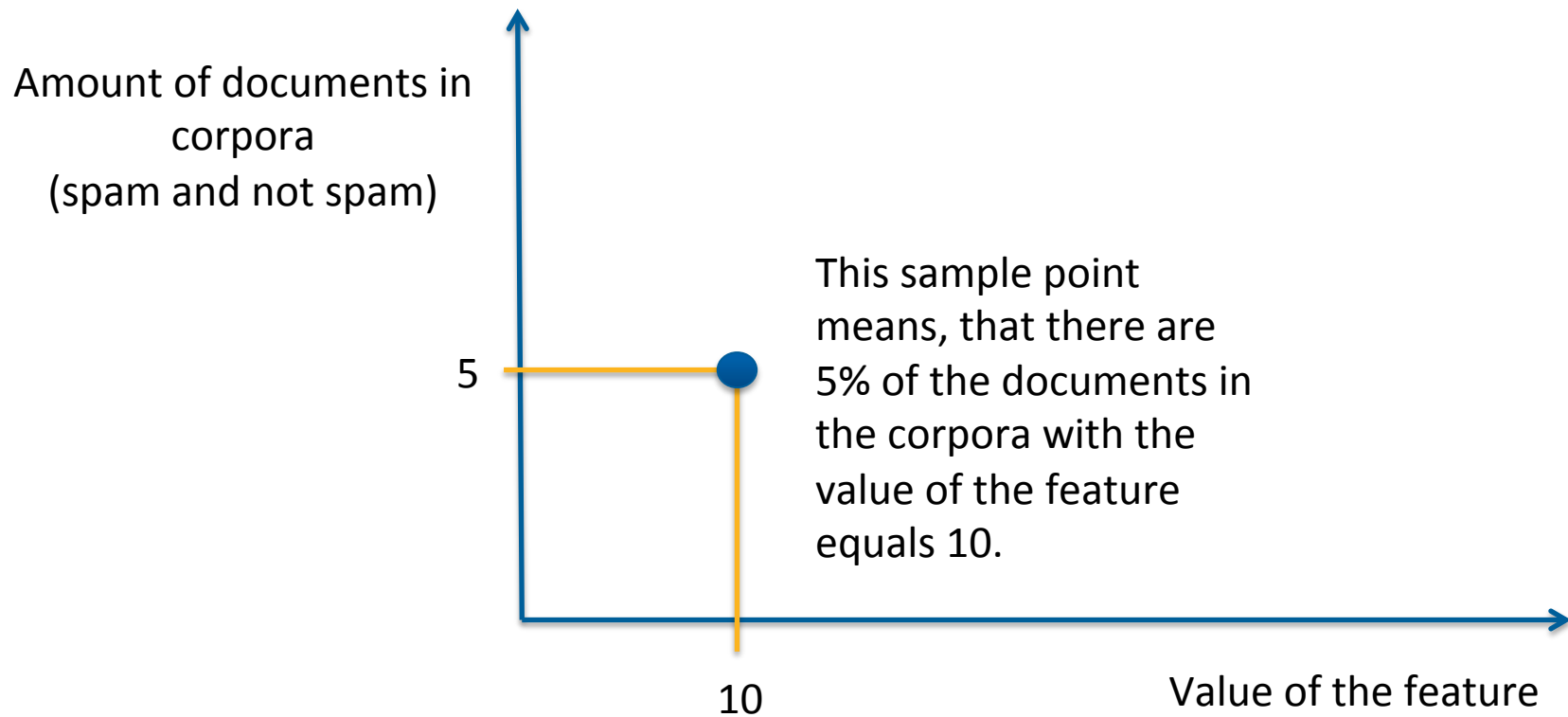
Classifier task

- We want to build classifier.
- Where a lot of machine learning methods.
- So we need:
 - Tagged corpora.
 - Features.
 - Evaluation metrics.
 - Some machine learning algorithm.
- I will talk about features.

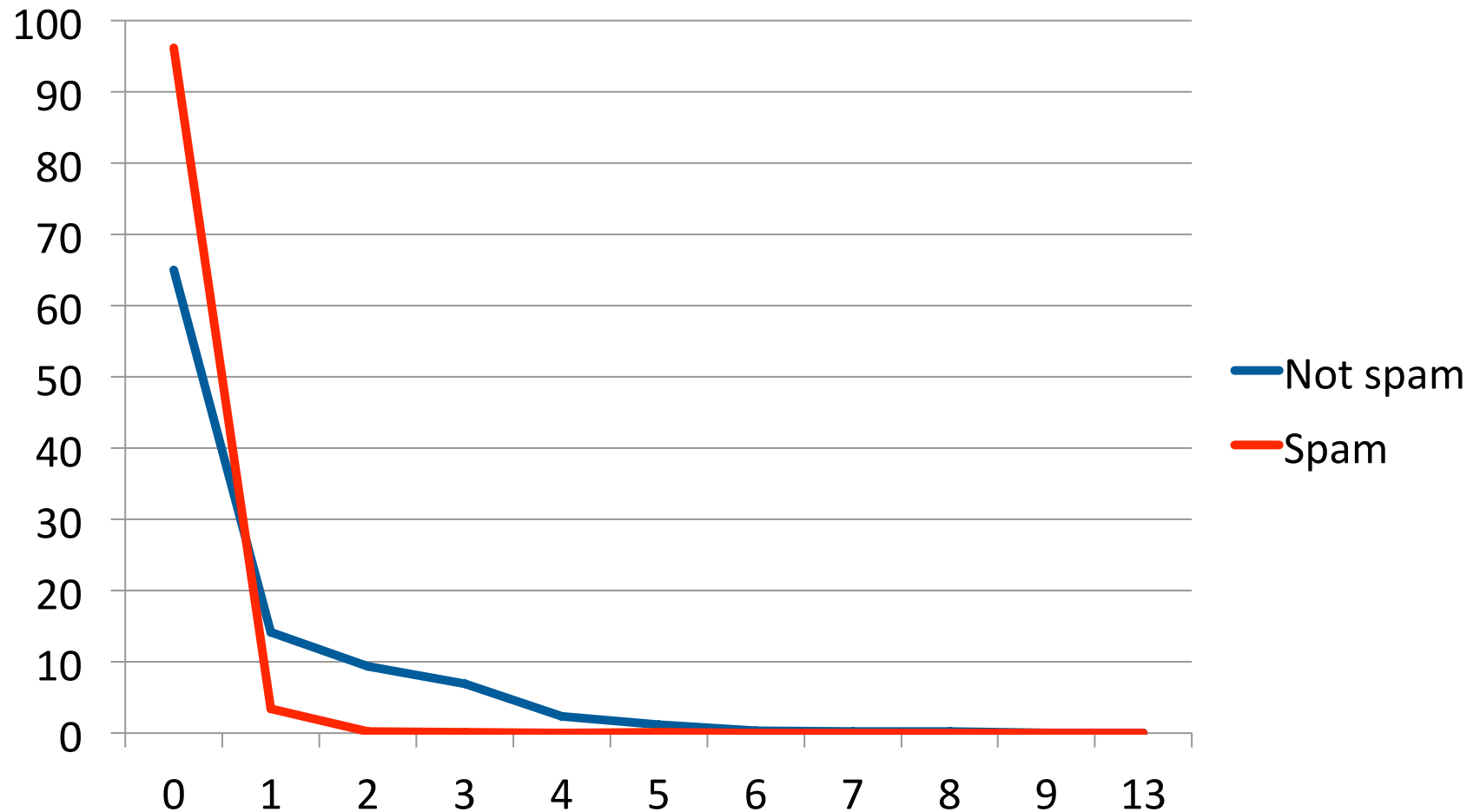
Web Page Features

- Only one page to analyze – no extra information
- Amount of ads
- Popular keywords
- HTML structure
- Language models

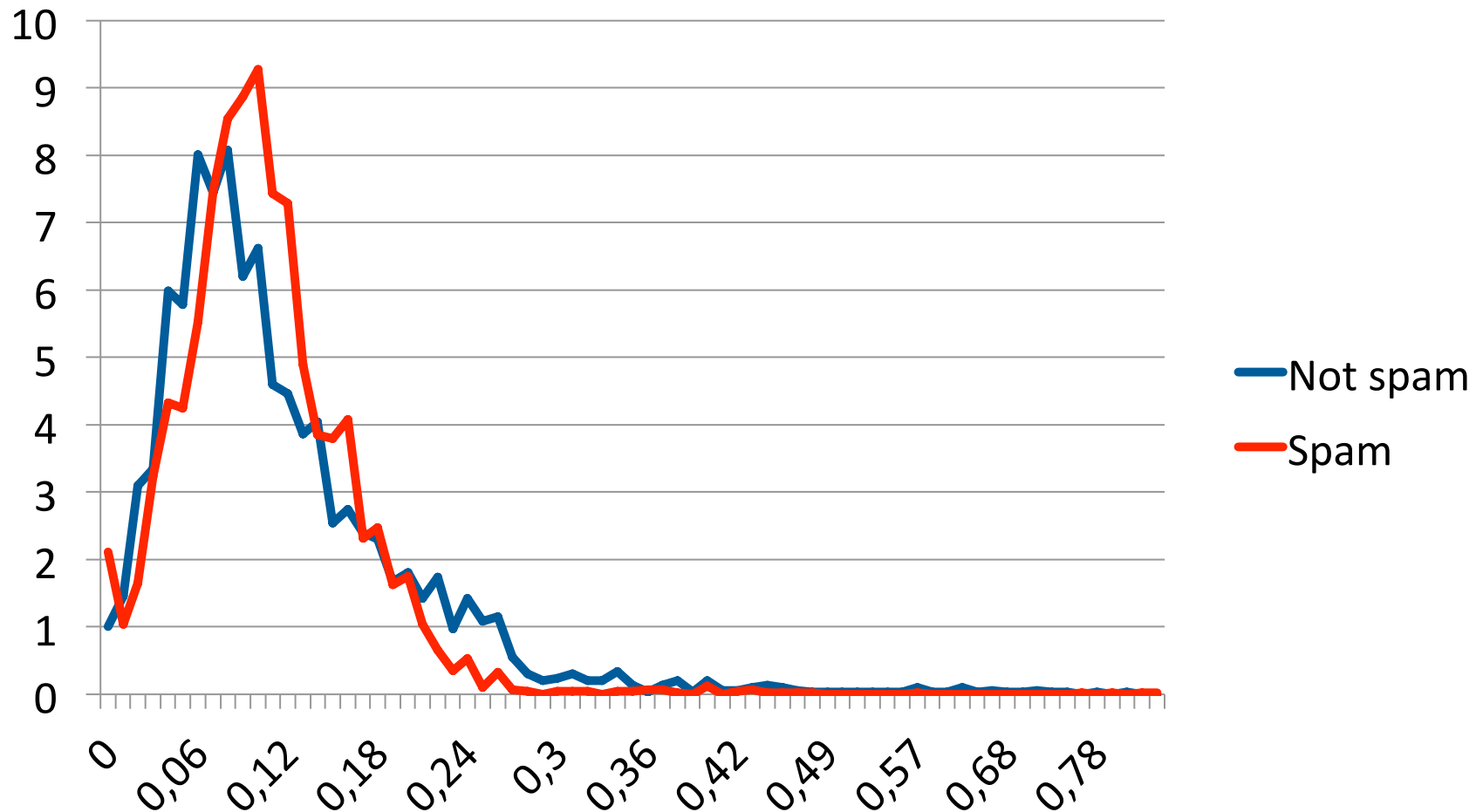
Format of graphs



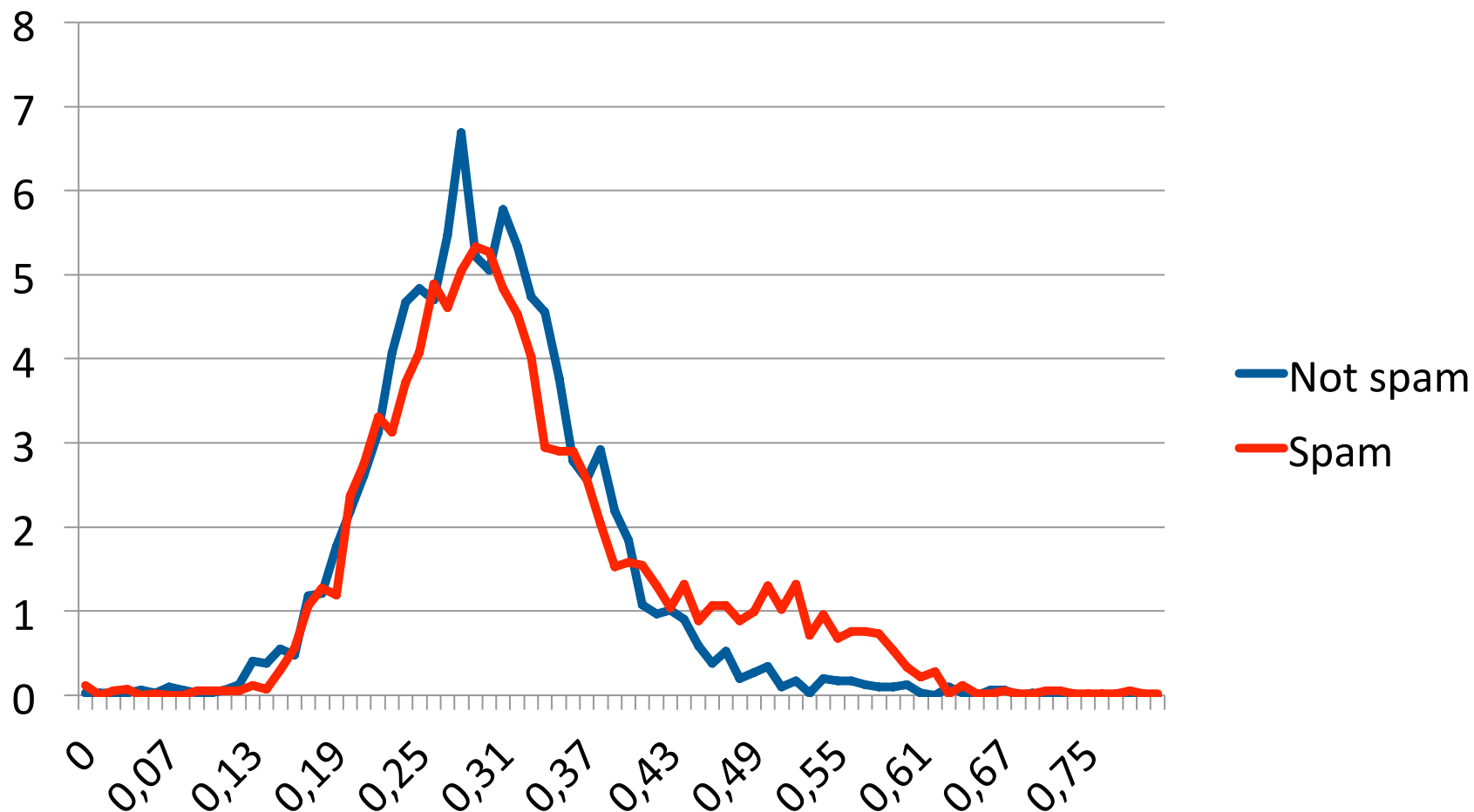
Amount of ads



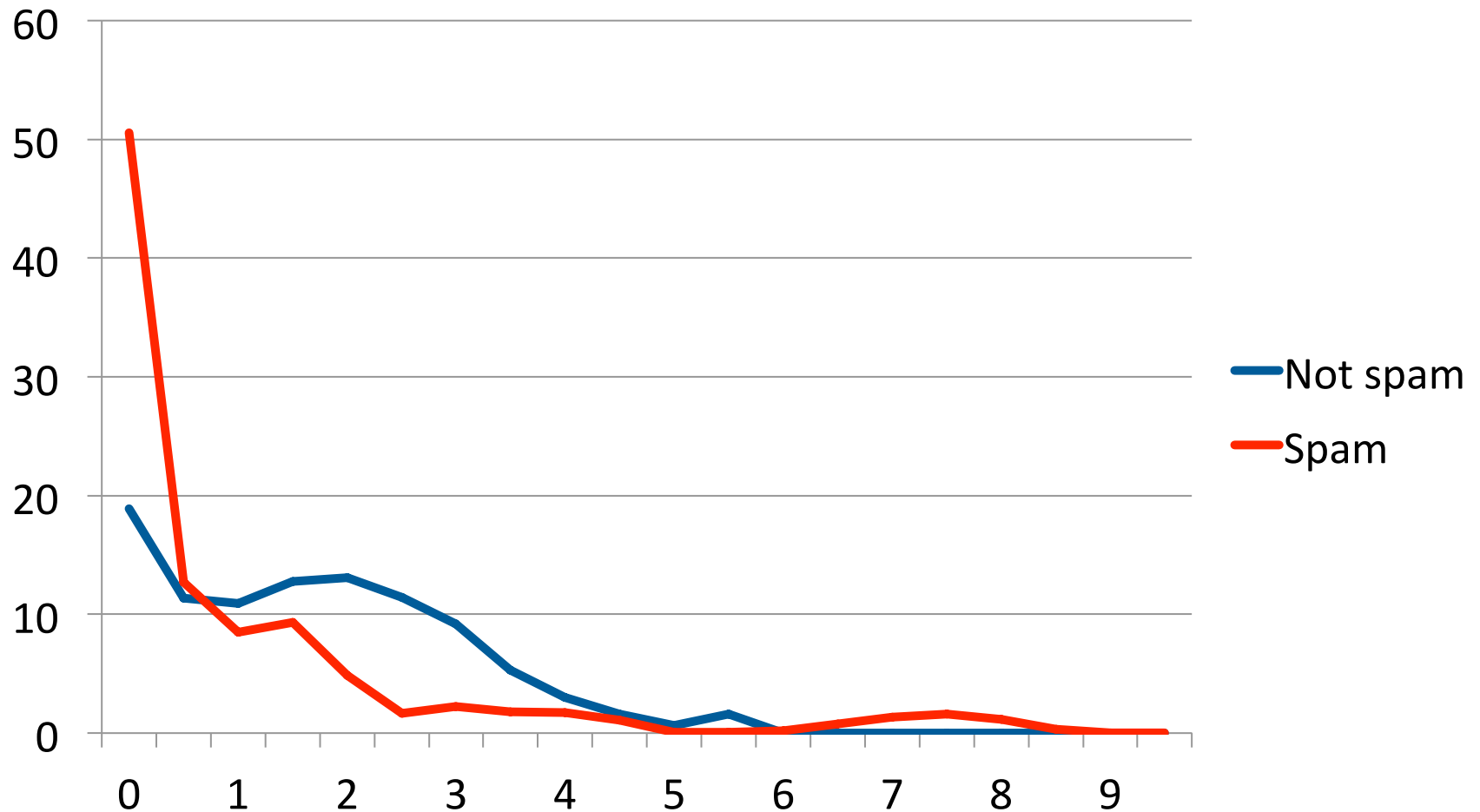
Proportion of top 500 keywords in terms from the document



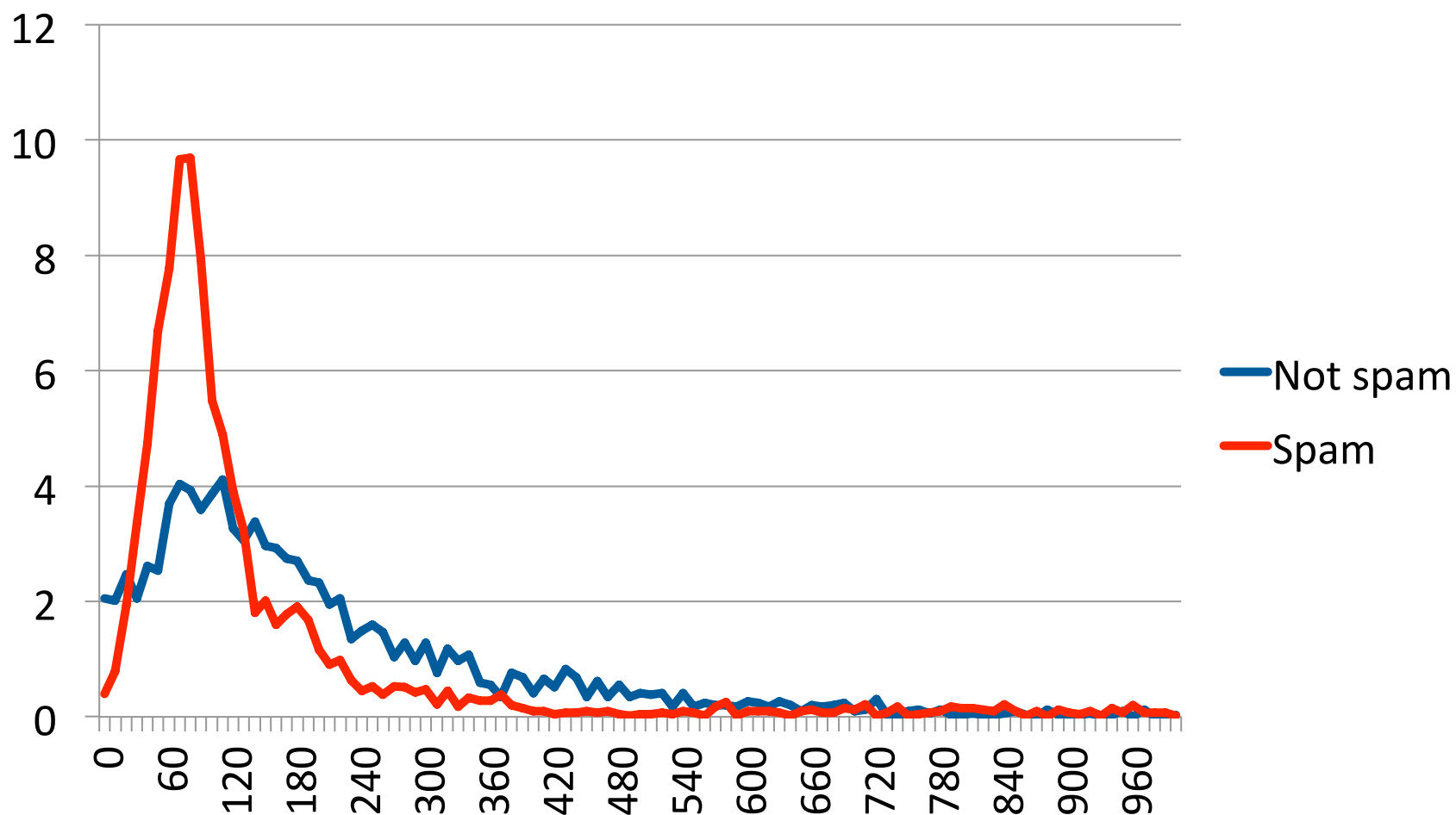
Proportion of terms from the document in the top 500 keywords



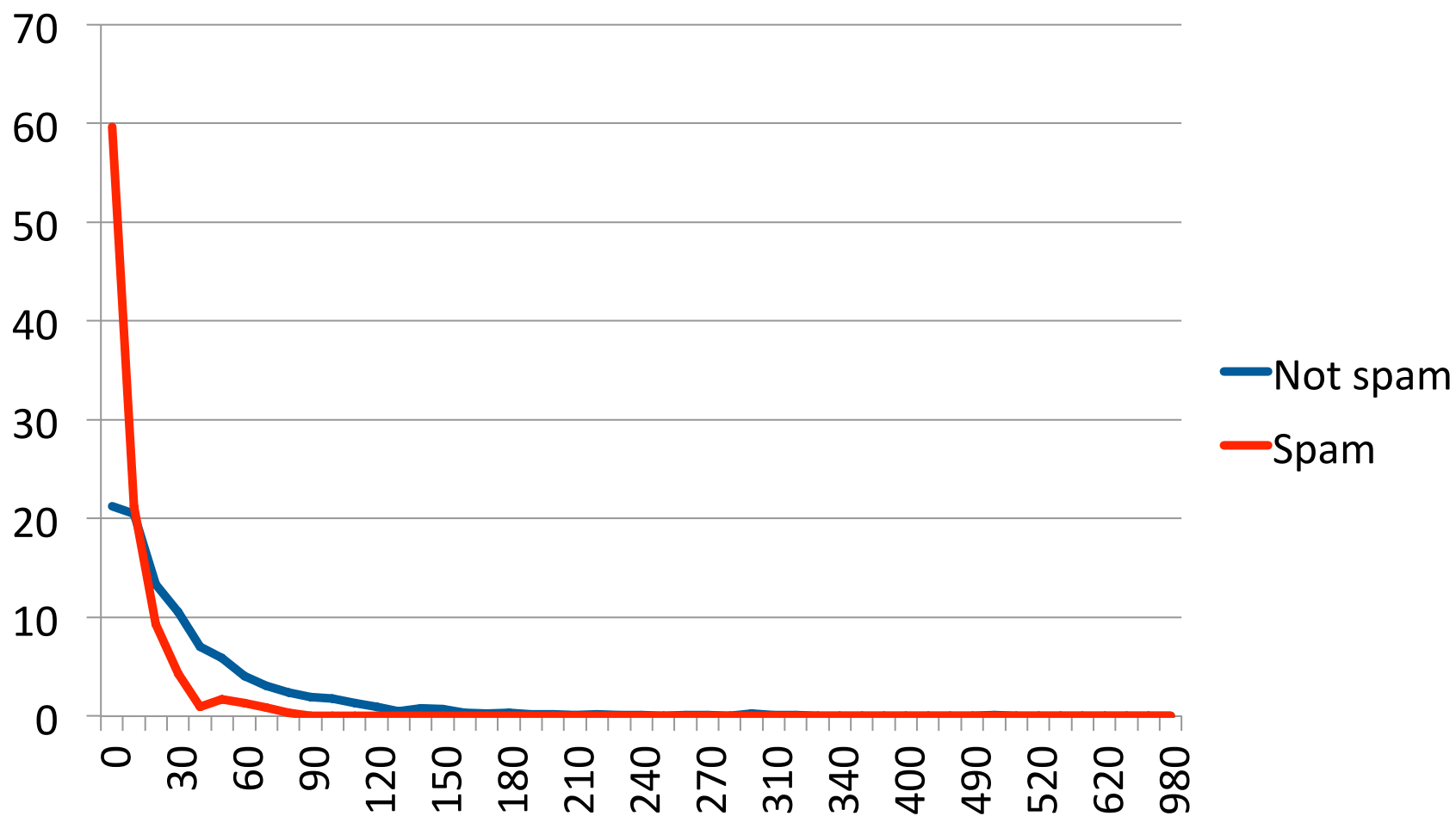
Entropy of host names in links (zero is bad)



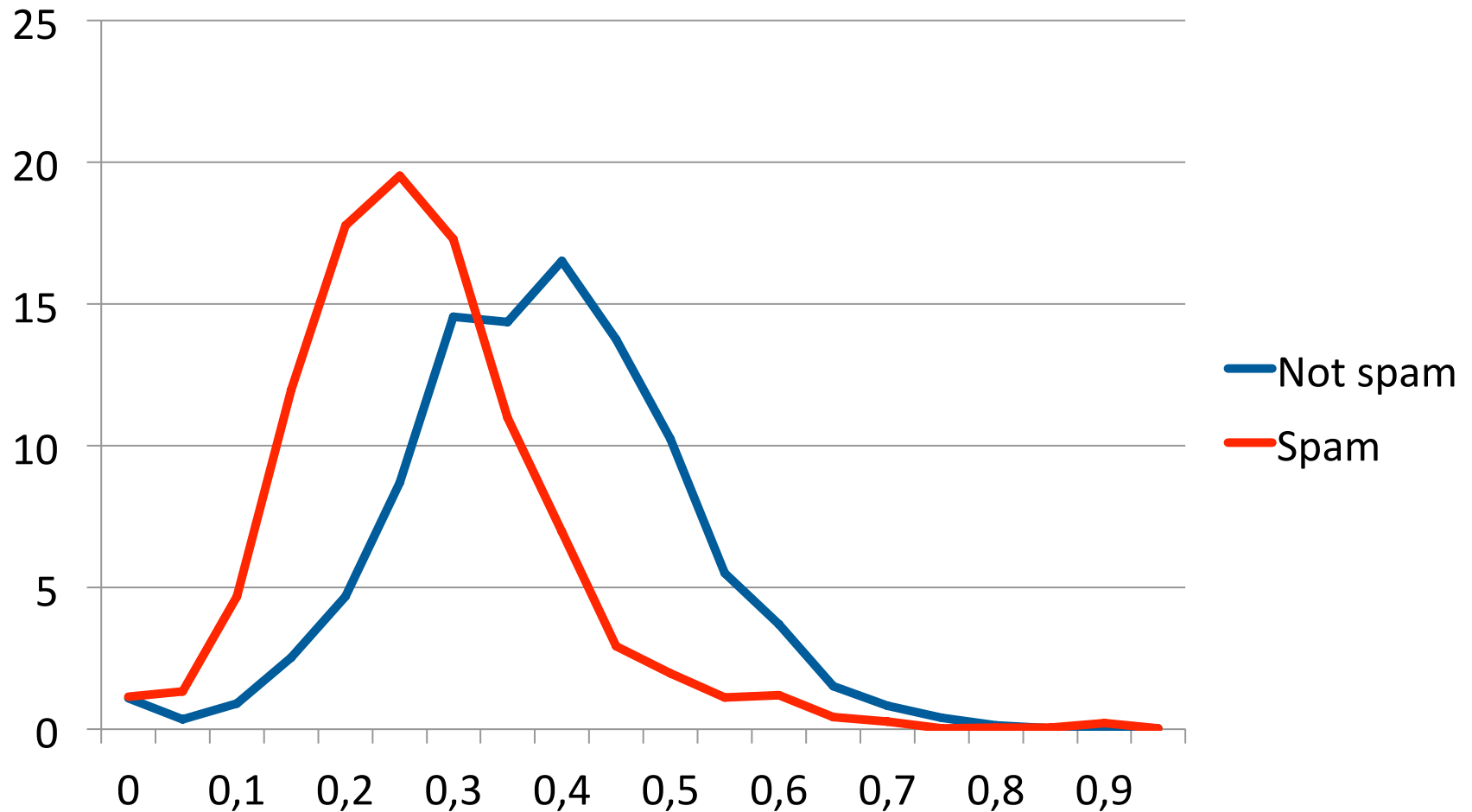
Amount of characters in link anchors



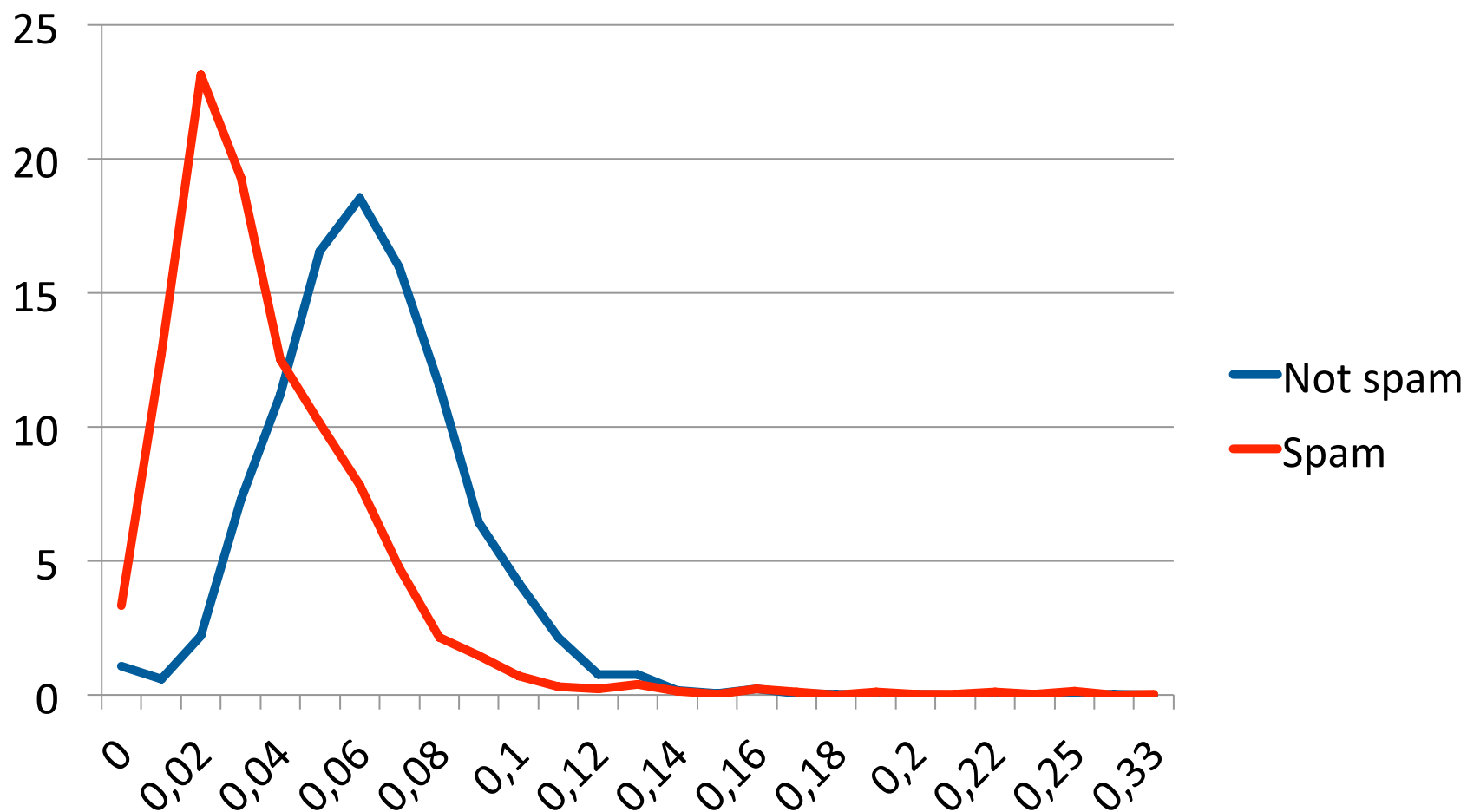
Number of images



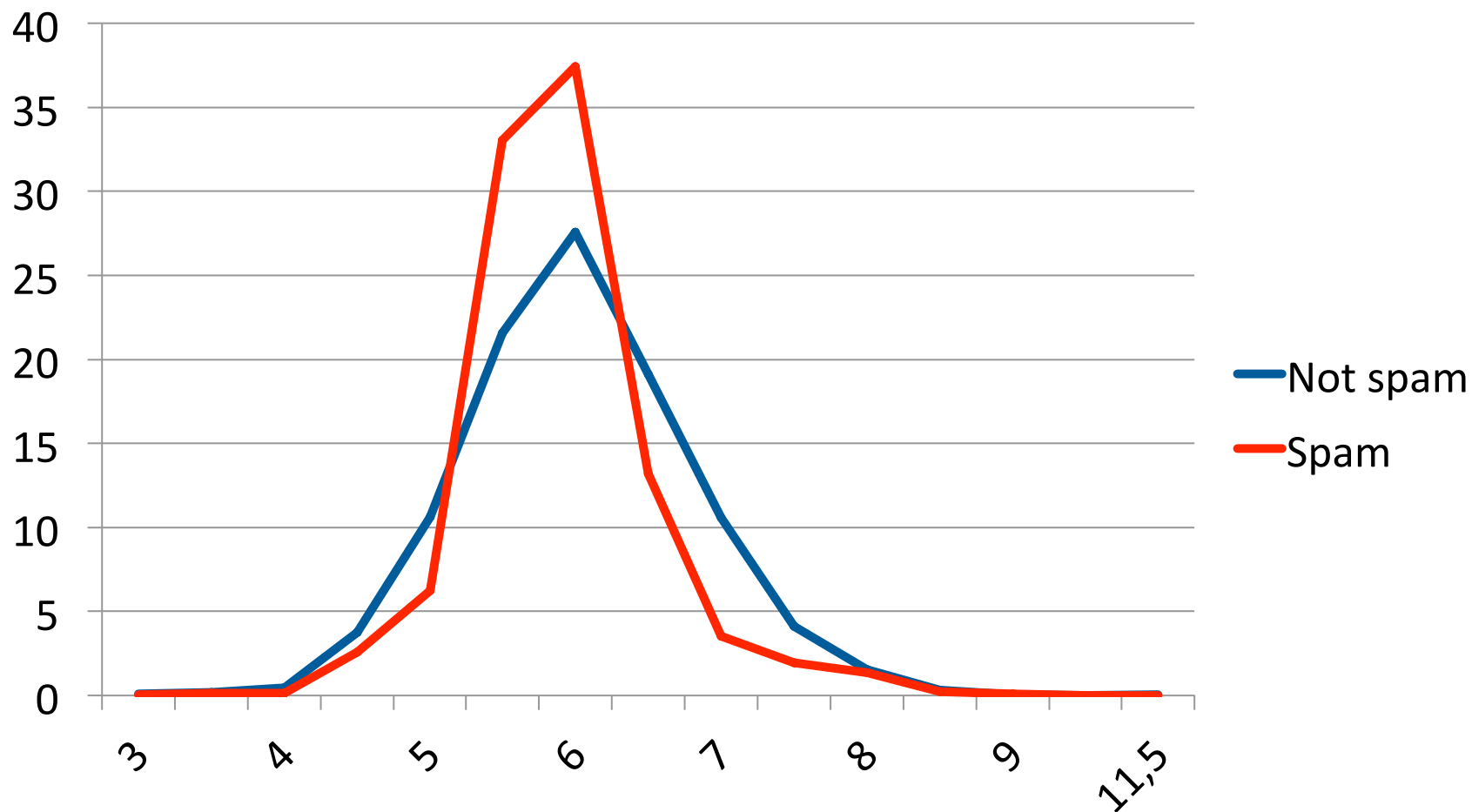
Average $tf*idf$ of important words



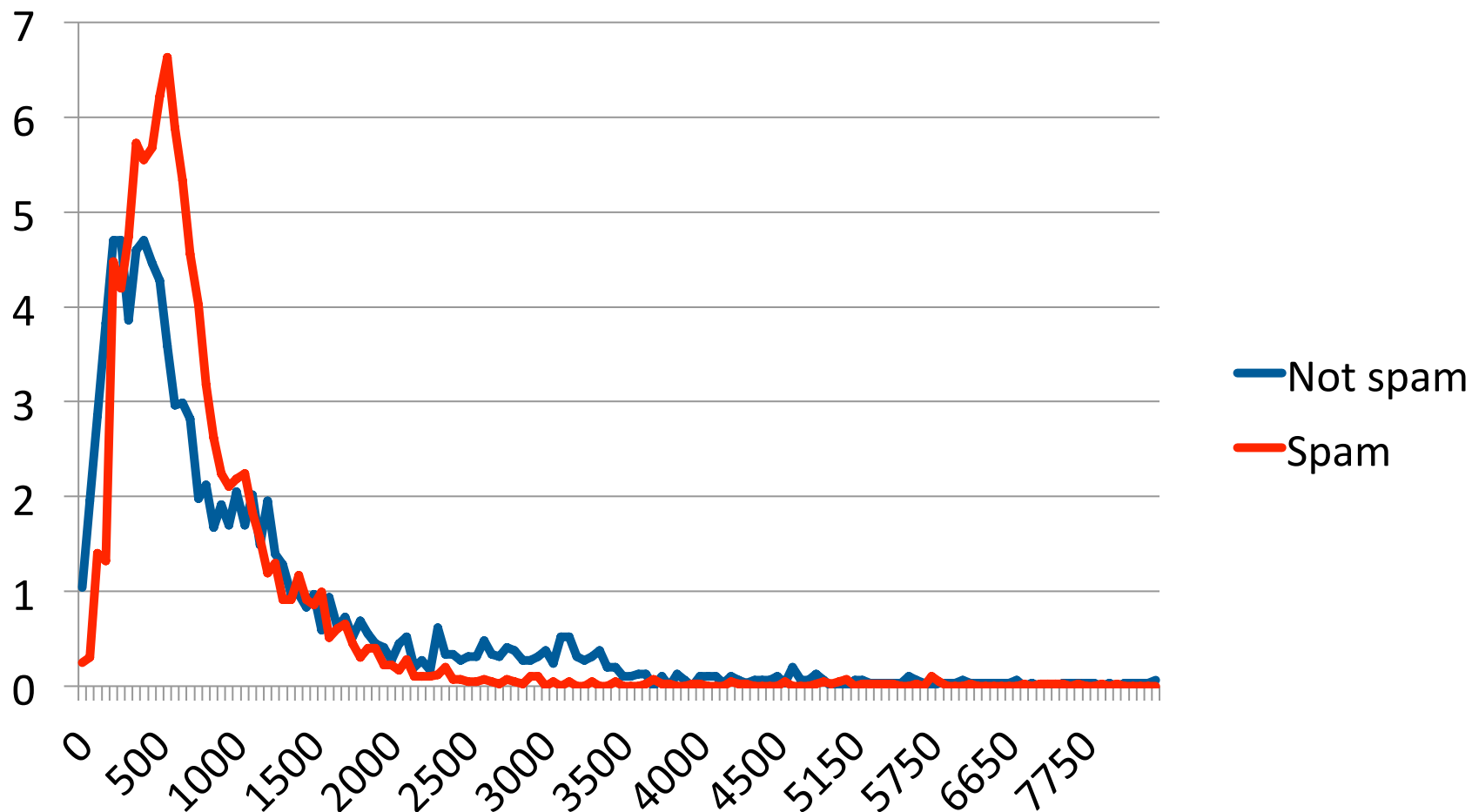
Proportion of important words



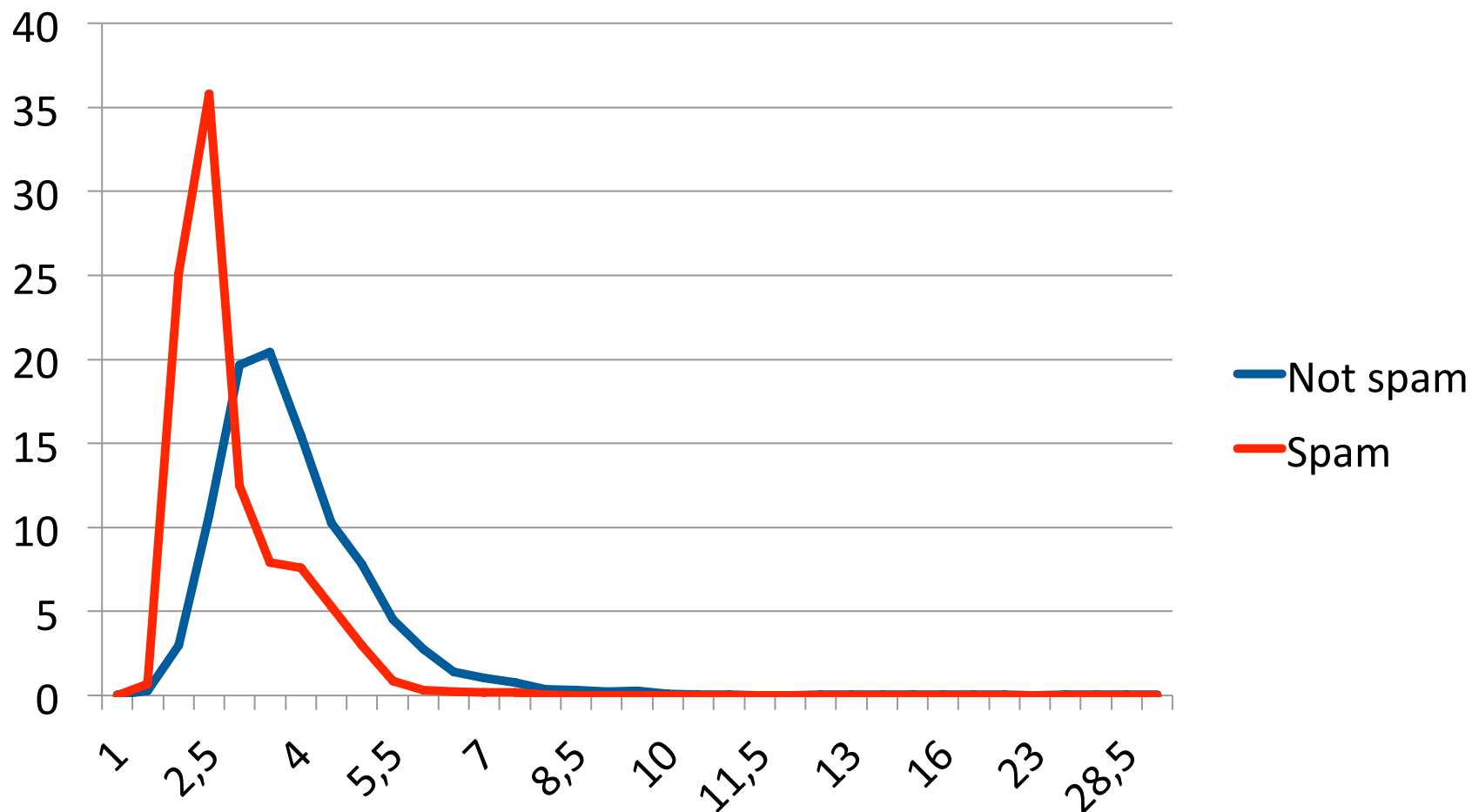
Average word length



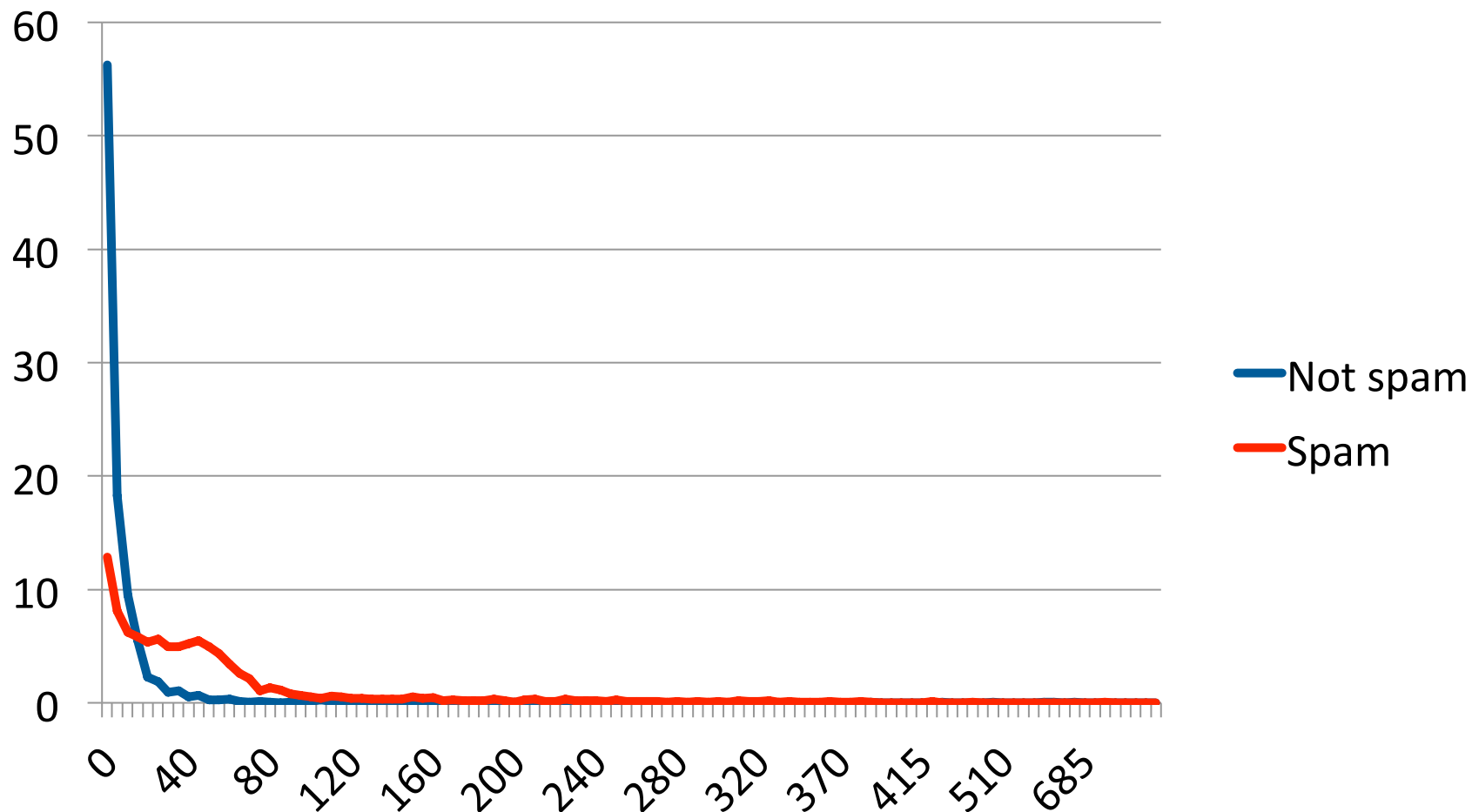
Document length (in words)



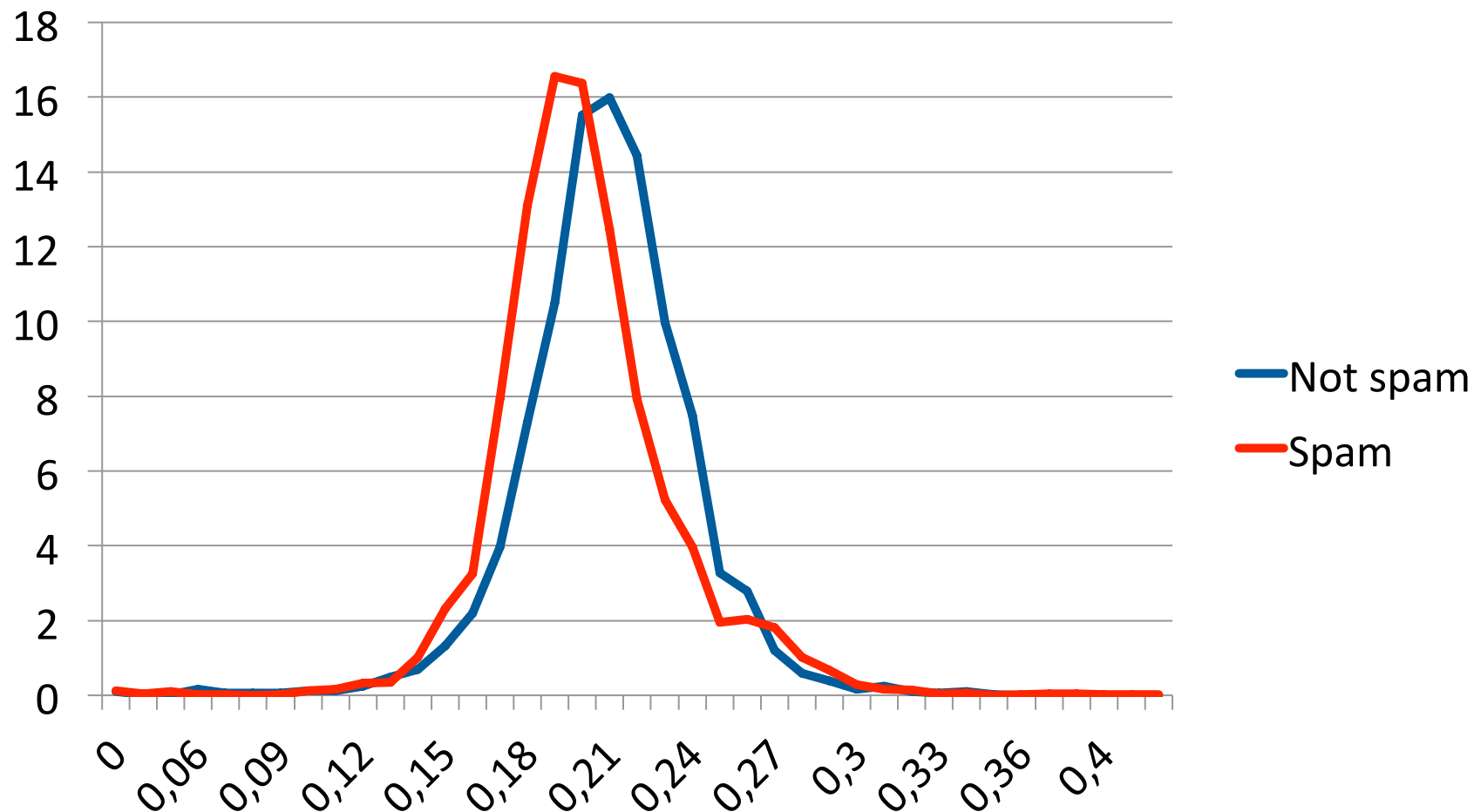
Compressibility



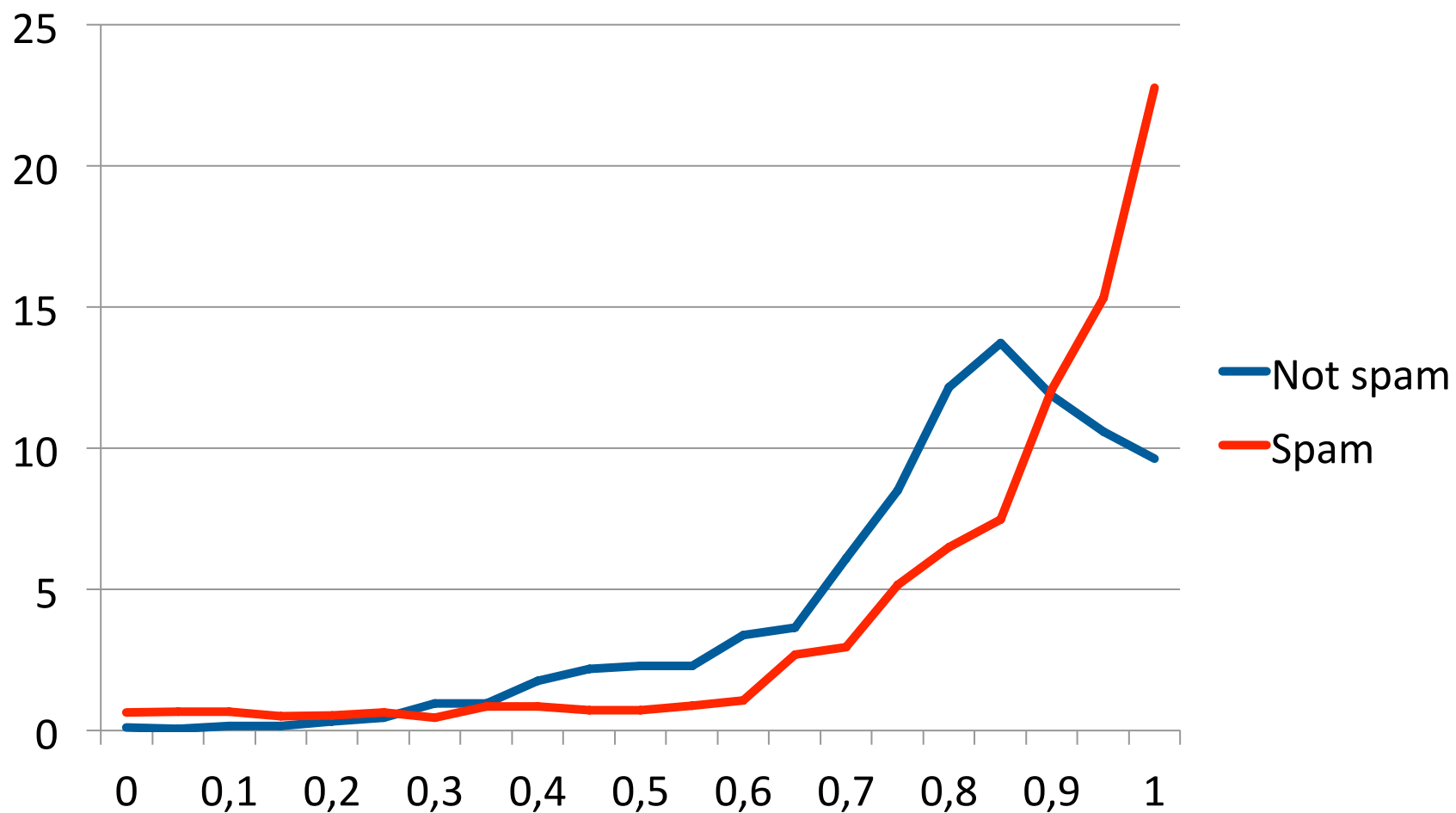
Popular spam words (free, download, etc...)



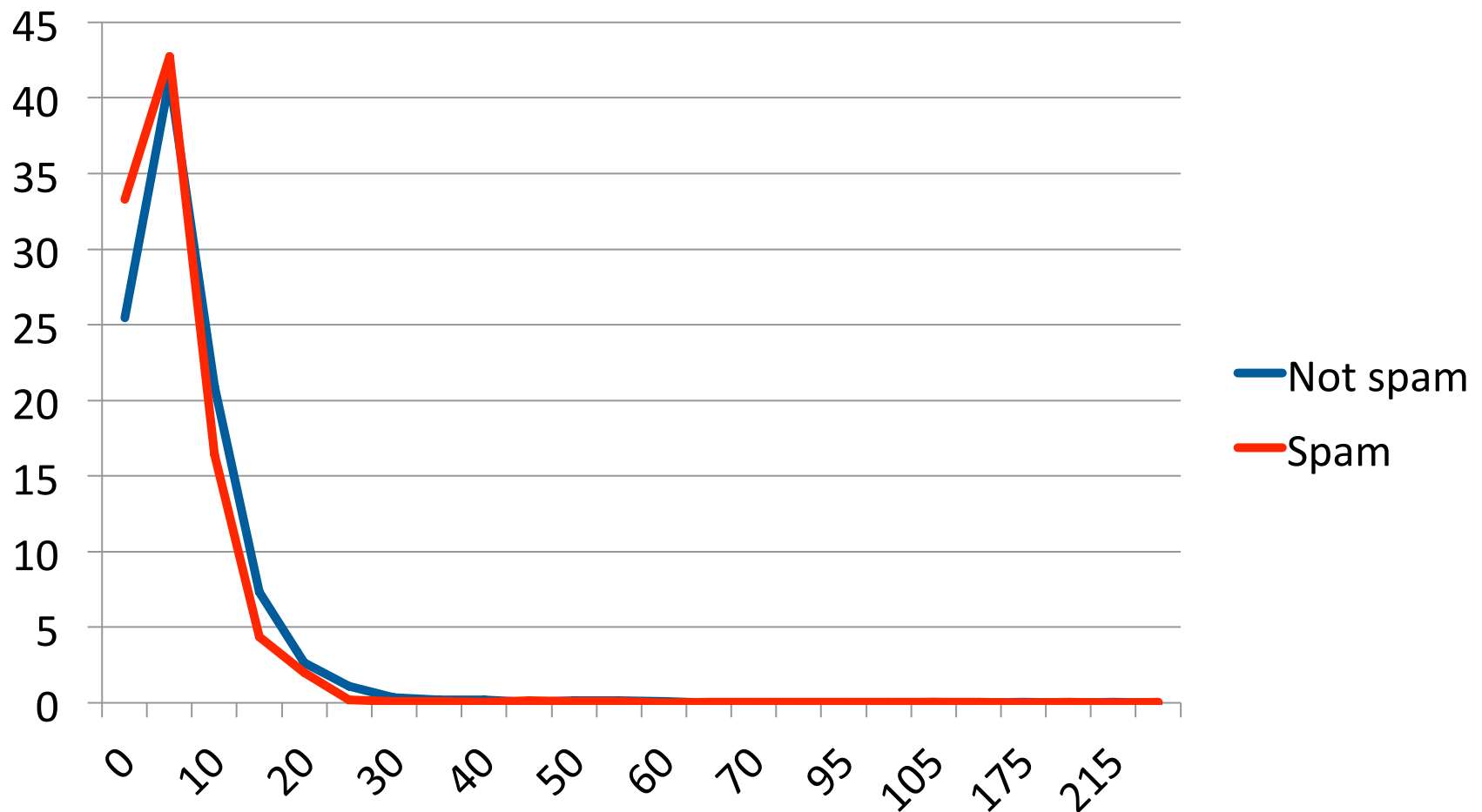
Language model (sum of logarithms)



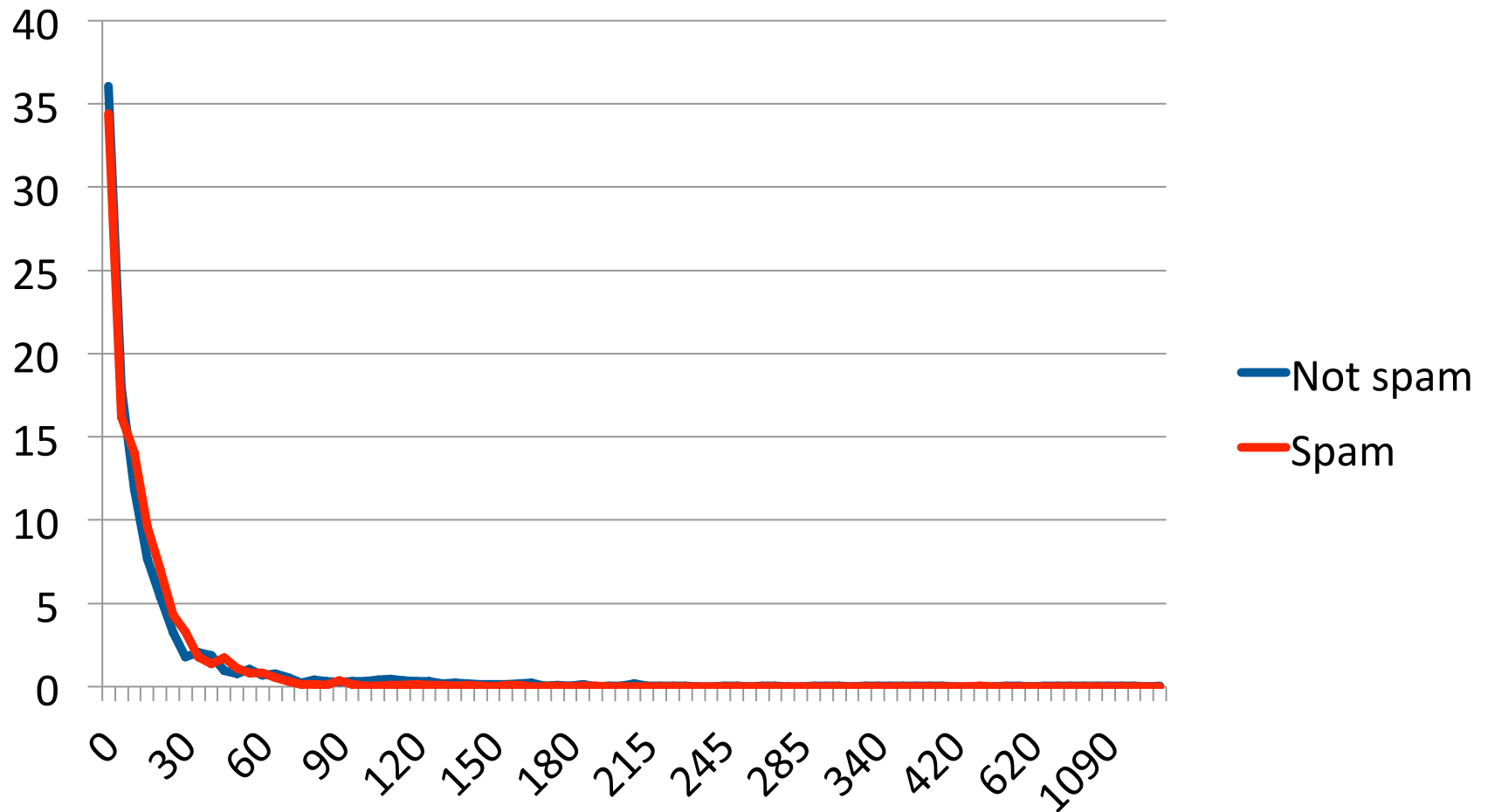
Amount of russian words



Number of words in title (no difference)



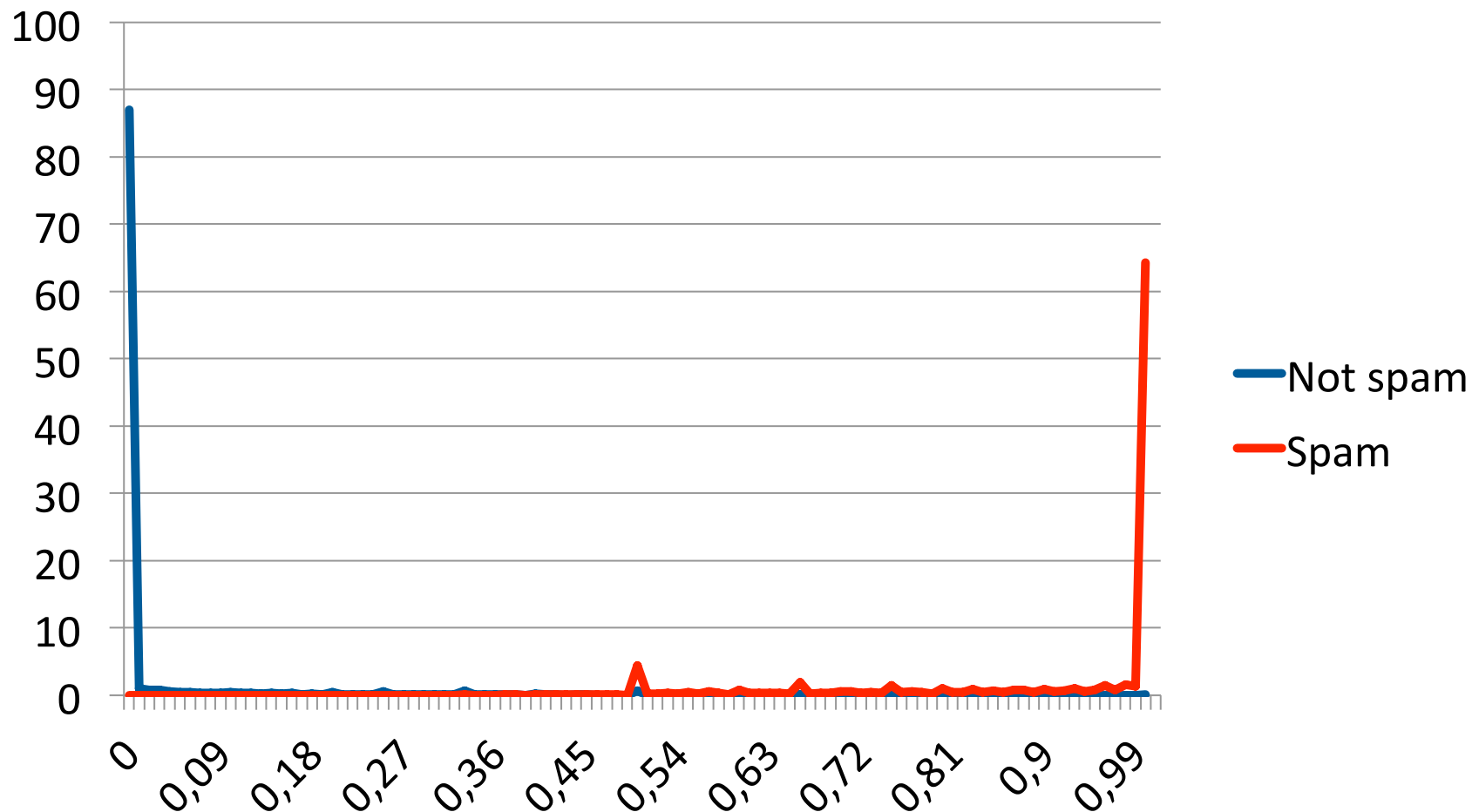
Number of words in html headers (no difference)



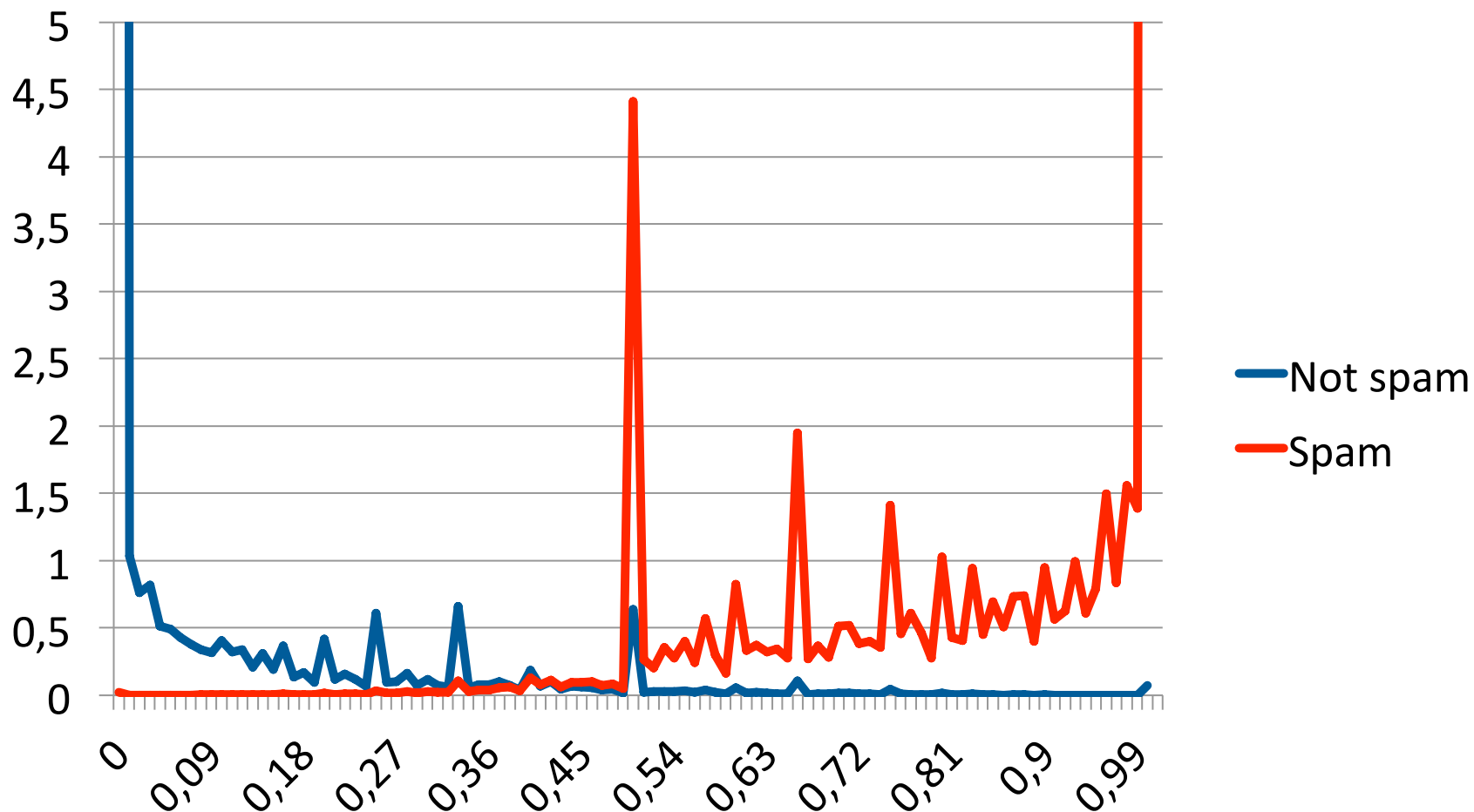
Just few slides about web site features

- Results of document classifier for each page.
- Link graph.
- Technical information.
- User behaviour.

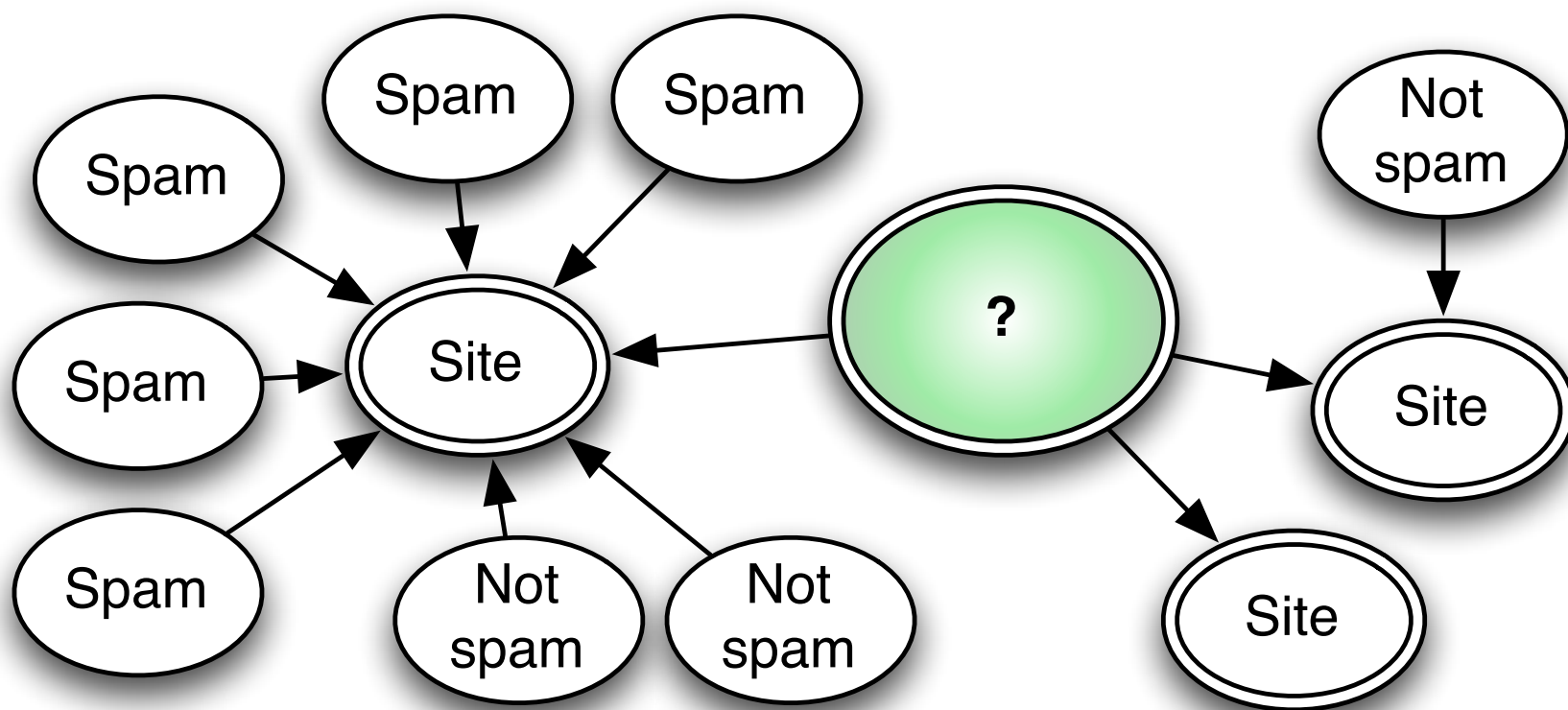
Amount of spam pages



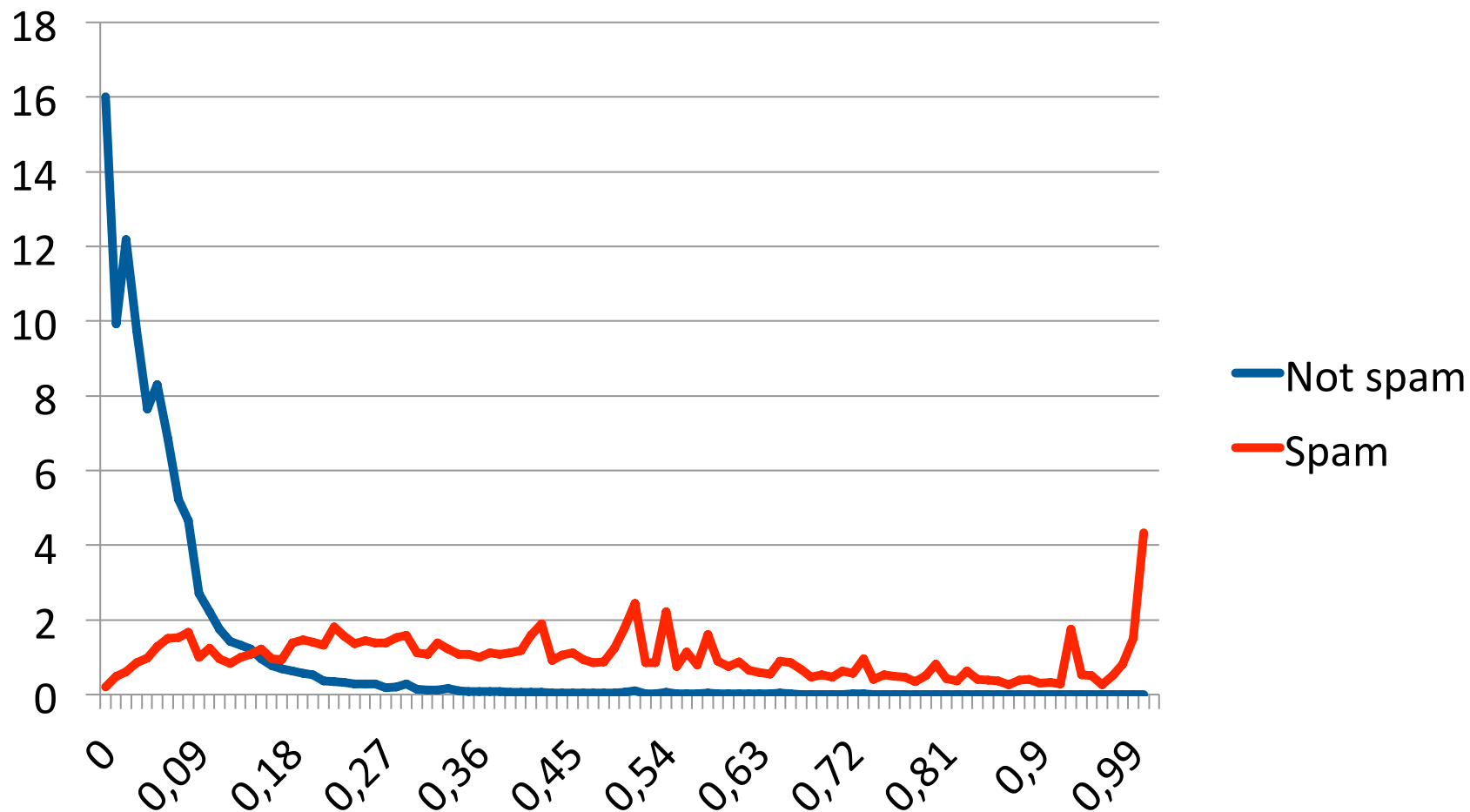
Amount of spam pages (scaled)



Link farm



Link farm probability



Spam classifier

- No “silver bullet” feature.
- We need to combine them.
- Document classifier: neural network.
- Site classifier: decision tree.
- For spam we have 87% precision and 80% recall rate.

Some extra notes about spam

- Spam is evolving.
- So, we can't build our classifier once.
- We need to constantly re-evaluate it.
- Spammers will throw away banned domains.
- So we have to recheck them sometimes.
- ... and most of new sites will not be much better.

поиск@mail.ru®

Thank you!

go.mail.ru

