1. **Технические заголовки**

Прочитайте технические заголовки сообщения и ответьте на вопросы:

* 1. От кого и кому было отправлено сообщение?
  2. Кто получил это сообщение?
  3. Напишите в хронологическом порядке список серверов, через которые прошло сообщение

Delivered-To: [kate1992@gmail.com](mailto:kate1992@gmail.com)  
Received: by [10.27.15.212](http://10.27.15.212/#_blank) with SMTP id 203csp366244wlp;  
 Mon, 18 May 2015 13:32:06 -0700 (PDT)  
X-Received: by [10.112.136.166](http://10.112.136.166/#_blank) with SMTP id qb6mr3809098lbb.54.1431981126716;  
 Mon, 18 May 2015 13:32:06 -0700 (PDT)  
Return-Path: <[test-2009@yandex.ru](mailto:test-2009@yandex.ru)>  
Received: from [forward2h.mail.yandex.net](http://forward2h.mail.yandex.net/#_blank) ([forward2h.mail.yandex.net](http://forward2h.mail.yandex.net/#_blank). [2a02:6b8:0:f05::2])  
 by [mx.google.com](http://mx.google.com/#_blank) with ESMTPS id dy9si7389450lac.58.2015.05.18.13.32.06  
 for <[kate1992@gmail.com](mailto:kate1992@gmail.com)>  
 (version=TLSv1.2 cipher=ECDHE-RSA-AES128-GCM-SHA256 bits=128/128);  
 Mon, 18 May 2015 13:32:06 -0700 (PDT)  
Received-SPF: pass ([google.com](http://google.com/#_blank): domain of [test-2009@yandex.ru](mailto:test-2009@yandex.ru) designates 2a02:6b8:0:f05::2 as  
 permitted sender) client-ip=2a02:6b8:0:f05::2;  
Authentication-Results: [mx.google.com](http://mx.google.com/#_blank);  
 spf=pass ([google.com](http://google.com/#_blank): domain of [test-2009@yandex.ru](mailto:test-2009@yandex.ru) designates 2a02:6b8:0:f05::2 as permitted  
 sender) [smtp.mail=test-2009@yandex.ru](mailto:smtp.mail%3Dtest-2009@yandex.ru);  
 dkim=pass [header.i=@linkedin.com](#_blank);  
 dmarc=pass (p=REJECT dis=NONE) header.from=[linkedin.com](http://linkedin.com/#_blank)  
Received: from [mxfront8h.mail.yandex.net](http://mxfront8h.mail.yandex.net/#_blank) ([mxfront8h.mail.yandex.net](http://mxfront8h.mail.yandex.net/#_blank) [[84.201.186.12](http://84.201.186.12/#_blank)]) by  
 [forward2h.mail.yandex.net](http://forward2h.mail.yandex.net/#_blank) (Yandex) with ESMTP id 107FF701FA1 for <[kate1992@gmail.com](mailto:kate1992@gmail.com)>; Mon,  
 18 May 2015 23:32:05 +0300 (MSK)  
Received: from [mxfront8h.mail.yandex.net](http://mxfront8h.mail.yandex.net/#_blank) ([[127.0.0.1](http://127.0.0.1/#_blank)]) by [mxfront8h.mail.yandex.net](http://mxfront8h.mail.yandex.net/#_blank) with LMTP id  
 GnMclzRg for <[test-2009@yandex.ru](mailto:test-2009@yandex.ru)>; Mon, 18 May 2015 23:32:05 +0300  
Received: from [maile-be.linkedin.com](http://maile-be.linkedin.com/#_blank) ([maile-be.linkedin.com](http://maile-be.linkedin.com/#_blank) [2620:109:c006:104::205]) by  
 [mxfront8h.mail.yandex.net](http://mxfront8h.mail.yandex.net/#_blank) (nwsmtp/Yandex) with ESMTPS id IOFXcNxOvX-W4O0UlGU; Mon, 18 May  
 2015 23:32:05 +0300 (using TLSv1.2 with cipher AES128-GCM-SHA256 (128/128 bits)) (Client  
 certificate not present)  
Authentication-Results: [mxfront8h.mail.yandex.net](http://mxfront8h.mail.yandex.net/#_blank); spf=pass ([mxfront8h.mail.yandex.net](http://mxfront8h.mail.yandex.net/#_blank): domain of  
 [bounce.linkedin.com](http://bounce.linkedin.com/#_blank) designates 2620:109:c006:104::205 as permitted sender)  
 smtp.mail[=m6qa2mssamgdpi587ivlhdfjlbovkj0oh3ix3z64escuar1sps174mr@bounce.linkedin.com](mailto:%3Dm6qa2mssamgdpi587ivlhdfjlbovkj0oh3ix3z64escuar1sps174mr@bounce.linkedin.com);  
 dkim=pass [header.i=@linkedin.com](#_blank)  
DKIM-Signature: v=1; a=rsa-sha256; c=relaxed/relaxed; d=[linkedin.com](http://linkedin.com/#_blank); s=proddkim1024;  
 t=1431981124; bh=vxGE/AijQ7o7t8VTTn2R1LxjphKflanE4odfPdTp4EI=;  
 h=From:Subject:MIMEVersion:Content-Type:To:Date:X-LinkedIn-Class:  
X-LinkedIn-Template:X-LinkedIn-fbl;  
 b=DPw9oUYON4fX+ImH6b/3E6I2F382sxAxe8gU49TjlaLS7MQnuWEBkrr6MkCbPyuIf  
 f025mONLddRjXBePlebgkfPOKWU7BtGNMf4+HFyOwdxXmuct9LotKKM47Zue2Y/qaZ  
 Fthx8L8lneRwDky3hVIW0sYV3hBAnsLsUylHeEv0=  
From: Arseniy Ivanov <[messages-noreply@linkedin.com](mailto:messages-noreply@linkedin.com#_blank)>  
Message-ID: <[1474297870.7221942.1431981124783.JavaMail.app@lva1-app2144.prod](mailto:1474297870.7221942.1431981124783.JavaMail.app@lva1-app2144.prod#_blank)>  
Subject: Мне бы хотелось добавить вас к сети моих профессиональных контактов в LinkedIn.  
MIME-Version: 1.0  
Content-Type: multipart/alternative; boundary="----=\_Part\_7221940\_1043636035.1431981124782"  
To: <[test-2009@yandex.ru](mailto:test-2009@yandex.ru)>  
Date: Mon, 18 May 2015 20:32:04 +0000 (UTC)

1. **Анализ спама**

В папке "письма" находится набор писем, нужно разложить их по 2м папкам "спам" и "не спам" и описать в текстовом файле, почему то или иное письмо спам или не спам.

1. **Консоль Linux**

В папке "рассылки" собраны две рассылки.

1) Нужно разложить их на две папки и перечислить использовавшиеся для этого команды на bash

2) Воспользовавшись вызовами команды grep с ключами, выведите:

a) список тем (subject) писем.

б) список урлов в формате:

http://tracking.msadcenter.msn.com/gid.gif?o=1

http://www.rolex.com.gripfork.com

http://www.rolex.com/images/email/BaselEmailWatch.jpg

т. е. без имён файлов и лишнего текста вокруг

**4. Регулярные выражения**

1) Задача спам-аналитика состоит в том, чтобы на основании нескольких примеров описать рассылку, но так, чтобы не пострадали легальные письма. Напишите регулярные выражения по темам (Subject) рассылок из предыдущего задания. Эти регулярные выражения должны описывать тему полностью (начинаться с ^ и заканчиваться $), однако срабатывать только на этих рассылках. То есть вариант типа Subject : ^.\*Rolex.com.\*$ не подходит, т.к. может сработать и на любой другой, в том числе и легальной рассылке.

2) Напишите регулярное выражение по строкам с телефоном,

регулярное выражение должно описывать именно этот номер

Телефон: 7 ( Ч95) 7ЧО~ 6 6~6 5

Телефон: 7 (Ч95)7 Ч0 -6 6 - 65

Телефон: 7( Ч 95 ) 7Ч 0 - 6 6~6 5

Телефон: 7(Ч 95)74 О -66 -6 5

Телефон: 7 (Ч 9 5 )7 Ч 0=66 -6 5

Телефон: 7 ( 49 5) 7 40 -6 6= 6 5

Телефон: 7 ( 4 95 ) 7Ч О - 6 6~6 5

Телефон: 7( 4 95 )7Ч O= 66~6 5

Телефон: 7( 4 9 5) 7 40 -6 6~6 5

Телефон: 7 ( Ч9 5) 74 O= 6 6=6 5

Телефон: 7 ( Ч 9 5 )7Ч 0~6 6~6 5

Телефон: 7 (495) 7ЧО -66 =6 5

Телефон: 7( 4 95 )7 4 0~ 66 -6 5

Телефон: 7 (4 9 5) 7 40=6 6= 65

Телефон: 7( Ч9 5 )7 ЧО -6 6 - 65

Телефон: 7 (Ч 95)7 40~ 66=65

Телефон: 7 ( Ч95 )7 ЧO=6 6= 65

Телефон: 7 (4 9 5) 74 O -66=6 5

Телефон: 7 (Ч 95 )7 ЧO - 66 ~65

Телефон: 7( 49 5)7 Ч O=6 6 - 65

Телефон: 7 ( Ч 9 5)7 Ч0 - 66=6 5

Телефон: 7( 49 5) 7 Ч0~66=6 5

Телефон: 7(Ч9 5 ) 74 О - 66~ 65

Телефон: 7( 49 5) 7 4O =6 6=65

Телефон: 7 ( Ч95) 7 Ч O=6 6= 65

Телефон: 7 ( Ч95) 7 ЧО~66 - 65

Телефон: 7 (Ч 9 5 )7 Ч О -6 6 ~6 5

Телефон: 7 ( 4 9 5 ) 7 ЧО~6 6 -65

Телефон: 7 (4 95 )7 4 0 - 6 6~6 5

Телефон: 7(Ч 9 5)7 4 O -66=6 5

Телефон: 7 (Ч9 5) 740 - 66 ~65

Телефон: 7( Ч9 5 ) 74 О~6 6 = 65

Телефон: 7( Ч95 ) 740=66 =6 5

Телефон: 7( Ч9 5 ) 7ЧO - 66 - 6 5

Телефон: 7 ( Ч9 5 ) 74 О -6 6~6 5

Телефон: 7(4 9 5 ) 7Ч 0=66=6 5

Телефон: 7 (Ч 9 5 )74 О=6 6~6 5

Телефон: 7 ( 49 5 ) 7 Ч0 = 6 6~ 65

Телефон: 7( Ч 9 5) 7 ЧО~6 6= 65

Телефон: 7( Ч95 ) 74 О~ 6 6 -6 5

Телефон: 7( Ч 95 )7 40 ~6 6~6 5

Телефон: 7( Ч95 ) 7 40 - 6 6 -6 5

Телефон: 7 (Ч95 )7 Ч О= 6 6~6 5

Телефон: 7( Ч 95 ) 74 О - 66 ~65

Телефон: 7( 49 5) 7 4О~6 6~6 5

Телефон: 7 ( Ч95 ) 7 4O=66~6 5

Телефон: 7( Ч 95 ) 74 0 ~66 -65

Телефон: 7( Ч 95 )7Ч O = 6 6~65

Телефон: 7 (Ч9 5 ) 7 4 0~ 66 -6 5

Телефон: 7 ( 495 ) 7 Ч0 -6 6~ 65

3) Упростите регулярки

Мы работаем с детекцией строк, т.е. нам важно, сработала регулярка или нет.

Например, строку Buy Viagra можно задетектировать любой из регулярок:

a. Buy Viagra

b. (Now and always )?Buy Viagra

c. Buy [V]ia(?:g)ra

d. .\*Buy Viagra.\*

e. Buy Viagra(Pi == 3.1415)?

Но самым простым и хорошим является вариант a.

Перепишите (упростите) следующие выражения так, чтобы они срабатывали на всём, что детектирует оригинальная регулярка, и не срабатывали на том, на чём она не работает.

a. (Are you )?[Ll]ooking for a (soulmate|bride|wife|girlfriend)?\? Maybe I'm the (?:woman|lady|girl) that (?:interest you|will be of interest to you)?

b. ^[^\W\D]{1,3}\_days\_of\_static/(\w+)|(\w+/\w+)$

c. ([\w\d]+\.){1,3}narod\.ru/index\.(exe|scr)?$

4) Верно ли следующее утверждение? Объясните, почему

[^\D\W] эквивалентно \d