

Integração de Dados

Extração de dados: Wrappers e Expressões Regulares

1

Recursos:

- Web:
 - <http://www.regular-expressions.info/>
 - <https://reulison.com.br/regex/>
 - <https://www.piazinho.com.br/>
 - <https://regexpr.com>
 - <https://regex101.com>
 - <https://regexpal.com>
 - <http://piazinho.com.br/ed4/exemplos.html#1>
 - <http://www.regextester.com/>
 - <http://myregexp.com/>

2

2

Extração de dados: motivação

- Grande quantidade de informação baseada em texto sem estrutura:
 - Páginas Web
 - Emails
 - *Log files*
 - Protocolos de comunicação
 - Ficheiros: .doc, .rtf, .xls, .txt, ...
- Como é **representada a informação**
 - Dados estruturado ou semiestruturados: Bases de dados, XML
 - Pesquisa de informação: SQL, XPath ou XQuery
 - Texto sem estrutura
 - Como fazer pesquisas?

3

3

Extração de dados: wrappers

- Para extrair dados de páginas web, fontes de dados semi-estruturadas ou sem estrutura:
 - Usar Wrappers
 - **podem ser construídos manualmente:**
 - Expressões regulares, XPath
 - **podem ser construídos automaticamente:**
 - aprendizagem automática

4

4

Expressões regulares: definição

- É um método formal de se especificar um **padrão** de texto.
- São usadas para procurar e/ou substituir texto que segue um determinado padrão ou possui determinadas restrições.
- É um conjunto de símbolos, caracteres literais e caracteres com funções especiais que, agrupados entre si, formam uma sequência, uma **expressão**. Essa expressão é interpretada como uma regra, que indicará sucesso se uma entrada de dados obedecer exatamente a todas as condições dessa regra.

5

5

Ou ainda...

- ER = uma forma de procurar um texto do qual não se possui toda a informação, mas apenas algumas variações possíveis (ex: *Pesquisar numa pauta o nome dos alunos que tenham o número de aluno começado por a2101*)
- ER = uma forma de um programador especificar padrões complexos que podem ser procurados e numa cadeia de caracteres (ex: *retirar de um log file todas as pessoas que acederam entre as 9:00 e as 12:00*)

6

6

Expressões Regulares: suporte

- As ER não são uma linguagem de programação!!
- As ER são suportadas e interpretadas por processadores REGEX
- As ER são suportadas por um grande número de linguagens de programação
 - Grep, egrep, textMate, Atom, Sublime, Notepad++, PHP,
 - RegexBuddy, RegexMagic
 - Perl, C, C++, Java, Javascript, Ruby, Python, .NET, Unix, Apache, MySQL

7

7

Expressões Regulares: versões

- POSIX (*IEEE Portable Operating System Interface, 1986*)
 - É um padrão mais “humano legível”
- PCRE (*Perl Compatible Regular Expressions*)
 - Padrão simplificado, usado por algumas linguagens
- Exemplo
 - ER para encontrar todos os dígitos de 0 a 9:
 - POSIX: **[0-9]**
 - PCRE: **\d**

8

8

Expressões Regulares

Usadas em inúmeras aplicações:

- Validação de entradas em formulários:
 - Formatos de endereços de email
 - Formatos de datas
 - Formatos de telemóveis ou telefones
- Pesquisa de palavras relacionadas ou “parecidas” em ficheiros
 - Center, Centre, Centered, ...
- Pesquisa de frases contendo uma parte variável:
 - **Eu vivo em** Coimbra // **Eu moro em** Coimbra // **Eu resido em** Coimbra
- Contar a frequência de palavras ou expressões numa frase
 - Eu **treino** todos os dias // o meu **treino** é diário // o **treino** de dança está marcado, o tutorial do **treino** está no Youtube, ...

9

9

Expressões Regulares

- Por exemplo:

Extrair dados de um cliente de um ficheiro

- Telemóvel
- Telefone fixo
- Código Postal
- Email
- Os que vivem em Coimbra
- Os que estão registados há mais de 10 anos,
-
- >> toda a informação que possa ser codificada num FORMATO / PADRÃO

10

10

Expressões Regulares

- Construídas por diferentes tipos de caracteres:
 - Literais
 - Metacaracteres:
 - representantes
 - quantificadores
 - âncoras
 - outros

11

11

Expressões Regulares: Literais

- Caracteres com correspondência literal
 - No entanto são case sensitive!
- >> existem *flags* em aplicações e linguagens de programação que eliminam a sensibilidade às maiúsculas e minúsculas.
- *Exemplo:*
 - 'amor' → "viva o amor" ✓
 - 'amor' → "a gata é amorosa" ✓
 - 'amor' → "o namoro acabou" ✓
 - 'amor' → "Amor é fogo que arde..." ✗

12

12

Expressões Regulares: Literais

- ER: ana



13

13

Expressões Regulares: Metacaracteres

- **Metacaracteres** são caracteres que mudam as ERs:
 - São caracteres com significados especiais
 - Permitem transformar caracteres literais em expressões muito versáteis

\ . * + ? - ^ \$ | [] {} ()

14

14

Expressões Regulares: Metacaracteres

- Metacaracteres:

- **Representantes**: representam um ou mais caracteres;
- **Quantificadores**: indicam o número de repetições permitidas de um carácter um bloco de caracteres;
- **Âncoras**: marcam uma posição específica na linha ou palavra;
- **Outros**: caracteres de escape, de agrupamento, etc;

15

15

Expressões Regulares: Metacaracteres

Representantes

- ponto **wildCard metacaracter**

Representa qualquer carácter exceto a mudança de linha (\n)

Exemplo: acrescentar o wildCard a uma expressão literal: **a.ma**

alma asma arma aLma aSma aAma a7ma aXma ✓

anima ✗
(dois caracteres **não são capturados** por um único wildCard)

16

16

Expressões Regulares: Metacaracteres Representantes

- ponto **wildCard metacaracter**

Outro exemplo : `.am.s`

>> encontra palavras com 5 caracteres, o 1º e o 4º podem ser qualquer caracter exceto o \n, os restantes caracteres são literais

camas
Lamas
FamAs
Vamos

17

17

Expressões Regulares: Metacaracteres Representantes

[] lista de caracteres permitidos

[^] lista de caracteres proibidos

Pode ser colocada em qualquer parte da ER, indicando o que é permitido (ou proibido) na substituição desse caracter:

[abc]lima >> no primeiro caracter só pode surgir 'a' ou 'b' ou 'c'
esta ER apenas captura estas três palavras: **alima blima clima**

[^abc]lima >> no primeiro caracter não pode surgir 'a' ou 'b' ou 'c'
esta ER apenas captura todas as palavras exceto estas: **alima blima clima**

18

18

Expressões Regulares: Metacaracteres Representantes

[] lista de caracteres permitidos

[^] lista de caracteres proibidos

Outros exemplos

[abc]lima[sS] >> alimas alimaS blimas blimaS climas climas

[aA]nit[aA] >> ???

A[^cd]B[cd]C[ef] >> ???

19

19

Expressões Regulares: Metacaracteres Representantes

• Listas de representantes mais genéricas

[a-z] um caracter minúsculo

[A-Z] um caracter maiúsculo

[a-zA-Z] um caracter minúsculo ou maiúsculo

[0-9] um dígito (0, 1, 2, 3, ...9)

[^a-z] um caracter que não seja minúsculo

[^0-9] um caracter que não seja um dígito

20

20

Expressões Regulares: Metacaracteres Representantes

- O caracter **-** indica um intervalo de caracteres seguidos na tabela ASCII

[a-d] um caracter minúsculo: **a** ou **b** ou **c** ou **d**
 [0-5] um dígito (0, 1, 2, 3, 4 ou 5)

[55-94]

Atenção!! O caracter **-** apenas cria intervalos entre caracteres ou dígitos! Logo: este intervalo **NÃO** apanha números entre 55 e 94.

Captura o **dígito 5**, o **intervalo de 5 a 9** e o **dígito 4**

[55-94]

O caracter **-** é um metacaracter, logo, quando se encontra dentro de uma lista, não é capturado como caracter literal.

Por exemplo

4-0 não é capturada por esta ER **[0-9][0-9]**

4-0 é capturada por esta ER **[0-9]-[0-9]**

21

21

Expressões Regulares: Metacaracteres Representantes

- Exemplos

[0-1][2-4][0-9]: números com três dígitos; 1º dígito pode ser 0 ou 1, 2º dígito pode ser o 2, o 3 ou o 4 e o último pode ser qualquer dígito.

123 **020** **139** **000 X** **0290 X**

[A-Z][a-z][a-z][0-9]: palavras com 4 caracteres; o 1º é um caracter maiúsculo, o 2º e o 3º são caracteres minúsculos e o último é um dígito

Ana1 **Fly7** **Zzz0** **aAa29 X** **Aço4 X**

[^BCD][a-z][^a-z] : palavras com 3 caracteres; o 1º é qualquer caracter exceto o A, B ou C; o 2º é um caractere minúsculo e o último é qualquer caracter exceto um caracter minúsculo.

An1 **Flo** **ZzZ** **Ba9 X** **An4t X**

22

22

Expressões Regulares: Metacaracteres

Representantes

• Exemplos

Kno[[^]tb]

Know
Knol
Kno.
Kno9
Kno;
Kno
Knot
Knob

[[^]0-9]A[a-c]

KAc
AAa
BAc
;Aa
!Ac
XAc
2Ac
8Aa

23

23

Expressões Regulares: Metacaracteres

• Exemplos

• [ab][a-zA-Z]

- a maria abriu a bela mala

• [abm][abe][[^]r]

- a maria abriu a bela mala

• [ab][ab][abr]

- a maria abriu a bela mala

24

24

Expressões Regulares: Metacaracteres

• Representantes:

• Exemplo 1

`[abcABC].[^A-Z]`

| | | | | | |
|-----|---|-----|---|-----|---|
| Ana | ✓ | Dar | ✗ | aXn | ✓ |
| Bom | ✓ | AnA | ✗ | Dna | ✗ |

• Exemplo 2

`.[abs][^0-9A-Z]`

| | | | | | |
|------|---|-----|---|-----|---|
| 1sss | ✗ | 5s5 | ✗ | 5sa | ✓ |
| abs | ✓ | xbs | ✓ | abS | ✗ |

25

25

Expressões Regulares: Metacaracteres

• Exemplos

`[a-zA-Z]id[0-9].`

| | | | | | |
|--------|---|-------|---|-------|---|
| AZid09 | ✗ | Zaz23 | ✗ | cid2 | ✗ |
| aid9r | ✓ | Zid3X | ✓ | cid22 | ✓ |

`Ana[bela].[123]`

| | | | | | |
|---------|---|------------|---|--------|---|
| Anabela | ✗ | Anabela123 | ✗ | Anabe1 | ✓ |
| Anae2 | ✗ | anabe3 | ✗ | Anala3 | ✓ |

26

26

Expressões Regulares: Metacaracteres

- **Quantificadores:**

Permitem aumentar o número de caracteres que pode aparecer numa posição

- Exemplo: 1**3333**2

- Como indicar que na 2ª posição o dígito **3** pode aparecer 4 vezes?

27

27

Expressões Regulares: Metacaracteres

- **Quantificadores:**

Permitem aumentar o número de caracteres que pode aparecer numa posição

| | | |
|-------|--------------|--------------------------|
| ? | interrogação | zero ou um |
| * | asterisco | zero, um ou mais |
| + | mais | um ou mais |
| {n,m} | chavetas | de n até m |
| {n} | chavetas | exatamente n |

28

28

Expressões Regulares: Metacaracteres

- **Quantificadores:**

São colocados após o caracter que se pretende quantificar

- Exemplo: 133332

- Como indicar que na 2ª posição o dígito 3 pode aparecer 4 vezes?

Resposta: 13{4}2

29

29

Expressões Regulares: Metacaracteres

- **Quantificadores: exemplos com ?**

- Exemplo: 1?20?

- O primeiro e o último dígitos são **opcionais**, logo esta ER valida:

120

20

12

2

30

30

Expressões Regulares: Metacaracteres

• Quantificadores: exemplos com *

• Exemplo: 1*20*

- O primeiro e o último dígito podem aparecer **zero ou mais vezes**, logo esta ER valida muitas combinações, apenas o dígito 2 é obrigatório:

| | | |
|-----|-----------|------|
| 120 | 111112000 | |
| 20 | 2000000 | ... |
| 12 | 111111112 | ... |
| 2 | | |

31

31

Expressões Regulares: Metacaracteres

• Quantificadores: exemplos com +

• Exemplo: 1+20+

- O primeiro e o último dígito podem aparecer **uma ou mais vezes**, logo esta ER valida muitas combinações, o dígito 2 é obrigatório uma vez, o dígito 1 e 0 têm de aparecer pelo menos uma vez:

| | | |
|--------|-----------|-----------------------|
| 120 | 111112000 | ... |
| 11120 | 12000000 | ... |
| 112000 | 111112000 | ... |
| 12 | 1112 200 | >> NÃO SÃO CAPTURADAS |

32

32

Expressões Regulares: Metacaracteres

• Quantificadores: exemplos com {}

• Exemplo: 1{2}20{0,3}

- O primeiro dígito aparece **duas vezes**, e o último dígito pode ser opcional ou aparecer **uma, duas ou três vezes**. Logo esta ER valida::

112 1120
11200 112000

12 11120 11200000 >> NÃO SÃO CAPTURADAS

33

33

Expressões Regulares: Metacaracteres

• Exemplo 1

a?b*c+d{1,4}

abbbbccdd ✓
ccccddd ✓

abcd ✓
acd ✓

aabcd ✗
abddd ✗

• Exemplo 2

[ab]?[cd]*[ef]{1,4}

ace ✓
Abcdeff ✗

bdffff ✓
acddccddcceef ✓

dddeee ✓
ccccee ✓

34

34

Expressões Regulares: Metacaracteres

Exemplos

| | | | | | |
|-------------|-------|--------|---------|---------|--|
| abs? | | | | | |
| ab ✓ | abs ✓ | as ✗ | bs ✗ | absss ✗ | |
| abs* | | | | | |
| ab ✓ | abs ✓ | abss ✓ | absss ✓ | bssss ✗ | |

| | | | | | |
|---------------------|--------|----------|----------|---------|------------|
| abs{2,5} | | | | | |
| abs ✗ | abss ✓ | absss ✓ | abssss ✓ | aaabs ✗ | abbbssss ✗ |
| a{2}b{3}s{4} | | | | | |
| aabbbssss ✓ | aabs ✗ | aabbss ✗ | absss ✗ | | |

35

35

Expressões Regulares: Metacaracteres

• Exemplos

[a-zA-Z]{2}[0-9]{2}.

| | | |
|----------|----------|---------|
| aa00xx ✗ | aa00xx ✗ | AZ99z ✓ |
| aa00 ✓ | aB123 ✓ | AZ99 ✓ |

golo{5}!+

| | |
|----------------|------------|
| golooooo!!!! ✓ | golooooo ✗ |
|----------------|------------|

golo{5}!*

| | |
|----------------|------------|
| golooooo!!!! ✓ | golooooo ✓ |
|----------------|------------|

36

36

Expressões Regulares: Metacaracteres

- **Âncoras: permitem controlar o início e o fim de uma ER**

| | | |
|----|-------------|--------------------------|
| ^ | circunflexo | início da linha |
| \$ | cifrão | fim da linha |
| \b | borda | início ou fim de palavra |

37

37

Expressões Regulares: Metacaracteres

- **Âncoras: permitem controlar o início e o fim de uma ER**

| | | |
|---|-------------|-----------------|
| ^ | circunflexo | início da linha |
|---|-------------|-----------------|

Exemplo: ER com o literal **Ana** aplicada neste texto:

Anabela Simões
 Ana Melo Costa
 Anastacia Ana
 Antonieta CAna

Exemplo: ER com o literal **^Ana** aplicada neste texto:

Anabela Simões
 Ana Melo Costa
 Anastacia Ana
 Antonieta CAna

38

38

Expressões Regulares: Metacaracteres

- **Âncoras:** permitem controlar o início e o fim de uma ER

\$ cifrão fim da linha

Exemplo: ER com o literal
Ana aplicada neste texto:

Anabela Simões
Ana Melo Costa
Anastacia Ana
Antonieta CAna

Exemplo: ER com o literal **Ana\$**
aplicada neste texto:

Anabela Simões
Ana Melo Costa
Anastacia Ana
Antonieta CAna

39

39

Expressões Regulares: Metacaracteres

- **Âncoras:** permitem controlar o início e o fim de uma ER

\b borda de palavra

Exemplo: ER com o
literal **\bAna**
aplicada neste texto:

Anabela Simões
Ana Melo Costa
Anastacia Ana
Antonieta CAna

Exemplo: ER com o
literal **Ana\b**
aplicada neste texto:

Anabela Simões
Ana Melo Costa
Anastacia Ana
Antonieta CAna

Exemplo: ER com o
literal **\bAna\b**
aplicada neste texto:

Anabela Simões
Ana Melo Costa
Anastacia Ana
Antonieta CAna

40

40

Expressões Regulares: Metacaracteres

• Outros:

- \ escape torna literal um metacaracter
- | ou alternativas de ERs
- () grupo delimita um grupo

41

41

Expressões Regulares: Metacaracteres

• O caracter de escape \

- Numa lista de representantes há 4 caracteres que precisam do caracter de escape caso seja necessário transformá-los em caracteres literais:] ^ \ -

42

42

Expressões Regulares: Metacaracteres

- O caracter de escape \

- Exemplo 1: ER para capturar as expressões:

```
int x[9];
int x{9};
int x(9);
```

tem de se usar \ para capturar o caracter] na segunda lista de representantes:

```
ER: int x[ [ { ( ] 9 [ \ ] } ) ] ;
```

43

43

Expressões Regulares: Metacaracteres

- O caracter de escape \

- Exemplo 2: ER para capturar as expressões:

```
program01
program-1
program\1
program_1
```

tem de se usar \ para capturar o caracter \ e o caracter - que estão na lista de representantes:

```
ER: program[0\ - _ \\]1
```

44

44

Expressões Regulares: Metacaracteres

- O caracter de escape \
- Fora de uma lista de representantes é necessário o caracter de \ para tornar os metacaracteres literais, por exemplo

>> ? * + .

exemplo 1: operações aritméticas com + e *

12+331 >> ER: [0-9]+\+[0-9]+

45*292 >> ER: [0-9]+*[0-9]+

45

45

Expressões Regulares: Metacaracteres

- O caracter |
- Pode ler-se “ou” e permite colocar várias alternativas numa ER

>> Ana|ana|anA

Muitas vezes é reescrito usando listas de representantes, que contêm as várias alternativas para uma posição

>> [Aa]n[Aa]

46

46

Expressões Regulares: Metacaracteres

- Grupos são criados com ()
- Permitem dividir uma ER em várias partes (vários grupos) para quantificações mais específicas ou para retirar apenas a informação de um desses grupos

Exemplo: qual o tipo de dados de um array?
 int x[9];
 char b[100];
 float y[55];

Grupo 1 -> permite retirar apenas esta informação da ER

ER: ([a-z]+) [a-z]\[[0-9]+\];

Qual o nome da variável? Pode criar-se um grupo 2:

ER: ([a-z]+) ([a-z])\[[0-9]+\];

47

47

Expressões Regulares: Metacaracteres

- Grupos: Exemplos

([AaBb]{1,6}):([0-9]{2}):(200[0-9])

- Identificar os grupos:

- Grupo 1 palavras começadas por A, a, B ou B e com tamanho máximo de 7 caracteres
- Grupo 2 identificador com dois dígitos
- Grupo 3 anos entre 2000 e 2009
- : literal delimitador

Anabela:02:2012

↑
grupo 1

↑
grupo 2

↑
grupo 3

48

48

Expressões Regulares: Metacaracteres

- Nome e ano das pessoas nascidas entre 1990 e 2000?

Ana Simões nasceu a 12-12-1995
 Rui Matos nasceu a 01-01-1987
 José Matias nasceu a 26-11-2002
 Daniel Silva nasceu a 10-05-2000
 Fernando Grave nasceu a 30-04-1997
 Margarida Moço nasceu a 01-01-1990

Grupo 1: Identifica o nome

Grupo 3: identifica o ano entre 1990 e 2000

`([a-zA-Zçõéã\s]*)` nasceu a `([0-9]{2}-[0-9]{2}-(199[0-9]|2000))`

49

49

Expressões Regulares: Metacaracteres

- Nomes das pessoas que entraram depois das 9:00?

06:13 – Tomé Lopes
 08:40 – Anabela Simões
 09:00 – Rui Lopes
 09:05 – Carlos Freire
 10:00 – Pedro Melo
 10:15 – Sofia Limões
 11:16 – Filipe Torres
 11:45 – Maria Ferro
 14:55 – Luís Paulo
 16:08 – Gustavo Matos
 22:14 – Maria Lurdes

Grupo 1: Identifica as horas

Grupo 4: nome das pessoas

`(09:[0-1-9][1-9][0-9])|([1-2][0-9]|1[0-9]|2[0-3]):[0-5][0-9] – ([a-zA-Zãöíéó\s]+)`

50

50

Expressões Regulares: Metacaracteres

- Diferenças entre () e []

| Expressão Regular | Padrões encontrados |
|-------------------|-----------------------------|
| (ab) (cd) | ab cd |
| [ab] [cd] | a b c d |
| a b c d | a b c d |
| [abcd] | a b c d |
| (abcd) | abcd |
| [ab]* | aaa bbb ababab aabbba |
| (ab)* | ab abab abababab |

51

51

Expressões Regulares: Metacaracteres

- Outros:

| | |
|----|--|
| \s | Qualquer caracter de espaçamento |
| \d | qualquer dígito -> <i>equivalente a [0-9]</i> |
| \D | qualquer caracter excepto dígito <i>equivalente a [^0-9]</i> |
| \n | new line (Windows) |
| \r | carriage return (Mac) |
| \t | tabulações |

52

52

Sugestões para construir uma ER

- Construa a ER por partes
- Comece por uma versão simplificada mas funcional
- Vá acrescentando os melhoramentos necessários passo a passo
- Exemplo:
 - Construa uma ER que valide datas no formato
 - dd/mm/aaaa ou dd-mm-aaaa

53

53

ER para validar uma data

ER1: [0-9]{2}/[0-9]{2}/[0-9]{4}
 ER2: [0-9]{2}[/-][0-9]{2}[/-][0-9]{4}
 ER3: [0123][0-9] [/-][0-9]{2}[/-][0-9]{4}
 ER4: [0123][0-9] [/-][01][0-9] [/-][0-9]{4}
 ER5: [0123][0-9] [/-][01][0-9] [/-][12][0-9]{3}
 ER6: ([012][0-9]|3[01]) [/-]([01][0-9])[/-]([12][0-9]{3})
 ER7: ([012][0-9]|3[01]) [/-](0[1-9]|1[012]) [/-]([12][0-9]{3})

54

54

EXPRESSÕES REGULARES

Exercícios

55

Exercícios:

- Escreva uma expressão regular que capture as palavras :
 - golo
 - golos
 - golo!
 - gooooolo!
 - gooolooooo!!!!
 - golos!
 - gooolooos!
 - goooooooloos!!!

56

56

Exercícios:

- **Construa uma expressão regular que encontre os nomes cujo primeiro nome comece em “An” e cujo apelido comece com “Si” ou com “Sa”**
 - Anabela Sintra
 - Antonieta Silva
 - Anselmo Sinatra
 - Aniceto Sala
 - Ana Santos

57

57

Exercícios:

- **Construa uma expressão regular que encontre todos os números inteiros num texto. Números começados por zero são inválidos? (1 12 2345 343234 1202)**
- **Construa uma ER que encontre todas as strings que contêm as 5 vogais, em qualquer número, mas por sempre por ordem alfabética (aeiou aaaeeiouuuu aaaeeeeiiiiioouuuu)**

58

58

Exercícios:

- **Construa uma expressão regular** que permita encontrar todos os números de telemóvel portugueses (iniciados por 91, 92, 93, 96 e com nove dígitos)
- Altere a expressão regular de forma a aceitar apenas números terminados em **00**.

59

59

Exercícios:

- **Construa uma expressão regular** que valide *passwords* com os seguintes requisitos
 - Começar por um carácter maiúsculo
 - Seguido de caracteres maiúsculos, minúsculos e/ou dígitos
 - Termine com um dígito
 - Tamanho entre 6 e 10 caracteres

60

60

Exercícios:

- **Construa uma expressão regular** que permita validar um e-mail que um utilizador introduz num formulário.
 começa por letra e pode ter dígitos, . ou _
 o domínio tem 2 ou 3 caracteres
 abs_1234@isec.pt
 abs@eden.dei.uc.pt
 abs.bekas_1972@yahoo.com
- Altere a expressão regular de forma a aceitar apenas emails do domínio **pt**

61

61

Exercícios:

- Escreva algumas cadeias binárias que sejam representadas pelas seguintes ERs:

$$\backslash b0(0|1)^*0\backslash b$$

$$\backslash b(01)^*\backslash b$$

$$\backslash b[01]^*\backslash b$$

$$\backslash b0+10^*10^*10+\backslash b$$

$$\backslash b(0|1)^*0(0|1)(0|1)\backslash b$$

62

62

Exercícios: Analise as ERs e indique quais as correspondências que estas encontram:

`^a(ab)*a$`

1. abababa
2. aaba
3. aabbba
4. aba
5. aabababa

`^ab+c?$`

1. abc
2. ac
3. abbb
4. bbc

`^a.[bc]+$`

1. abc
2. abbbbbbb
3. azc
4. abcbcbcb
5. ac
6. asccbbbbcbbccc

`^(abc|xyz){2}$`

1. abcxyzaa
2. abcxy
3. abcd
4. abc|xyz.
5. xyzww

63

63

Exercícios: Analise as ERs e indique quais as correspondências que estas encontram:

`[a-z]+[\.\?!]`

1. battle!
2. Hot
3. green
4. swamping.
5. jump up.
6. undulate?
7. is.?

`[a-zA-Z]*[^\,]=`

1. Butt=
2. BotHEr,=
3. Ample
4. FIdDIE7h=
5. Brittle =
6. Other.=

`[a-z][.?!]\s+[A-Z]`

(\s == caracter de espaçamento)

1. A. B
2. c! d
3. e f
4. g. H
5. i? J
6. k L

`<[^>]+>`

1. <an xml tag>
2. <opentag> <closetag>
3. </closetag>
4. <>
5. <abs>
6. <with attribute="77">

64

64

Exercícios:

- Analise a seguinte ER e assinale quais as sequências encontradas:

1{3}0?(1*|2{2})

- A - 130111
- B - 111
- C - 1110111
- D - 130?1*
- E - 11122
- F - 1110122
- G - 111022

65

65

Exercícios:

- Analise a seguinte ER e escreva algumas frases reconhecidas pela ER

^[aAoO]\s.{3,5}\sestá\s((ap)|(rep))(rovad)([ao])\$

66

66

Exercícios:

- Analise a seguinte ER e indique quais as correspondências que esta encontra:

(very)+(fat)?(tall |ugly) man

- 1) very fat man
- 2) fat tall man
- 3) very very fat ugly man
- 4) very very very tall man

67

67

Exercícios

- Expressão regular que encontre frases começadas por vogais maiúsculas e terminadas por ponto final. Assuma que na frase podem surgir apenas os caracteres alfabéticos maiúsculos e/ou minúsculos e espaços em branco.

68

68

Exercícios:

- Escreva uma expressão regular que encontre frases interrogativas de tamanho entre 10 e 20 caracteres. Nas frases assuma que começam por qualquer caracter maiúsculo, seguidos de caracteres minúsculos, espaçamentos ou dígitos e terminando com o sinal de interrogação.

69

69

Exercícios:

- Escreva uma expressão regular que encontre as palavras que comecem por **ga**, **go**, **gi** ou **gr** seguidas de qualquer caracter minúsculo.
- Na frase seguinte, as palavras a sublinhado mostram exemplos do que a ER deve validar.

o gato preto e grande gosta de gaivotas brancas gigantes

70

70

Exercícios:

- Imagine que possui o seguinte ficheiro de coordenadas. Escreva a ER que capture quais as cidades cuja **Latitude** se situa a **N** e **Longitude** a **W**. Indique os grupos relevantes.

```
41°9'N 8°38'W Porto
38°42'N 9°11'W Lisboa
51°30'25"N 0°07'39"W Londres
22°54'30"S 43°11'47"W Rio
55°45'8"N 37°37'56"E Moscovo
```

71

71

Exercícios:

```
Sofia Melo *** CR *** 15 valores
Pedro Mota *** ID *** 6 valores
Rui Matos *** IIA *** 1 valor
Carlos Lopes *** P00 *** 18 valores
Sandra Serra *** IP *** 7 valores
Romeu Torres *** P00 *** 20 valores
```

- Recorrendo a grupos, construa uma expressão regular que encontre os nomes das pessoas, respetivas disciplinas e classificações, mas apenas nas disciplinas com notas superiores ou iguais a 15 valores. Indique os grupos necessários de forma a retirar o **nome**, a **disciplina** e a **nota** que respeitem o critério acima definido.

72

72

Exercícios:

Caramelo - gato silvestre - vacinado em Coimbra
 Tareco - gato domestico - vacinado em Antanhol, Coimbra
 Calvin - cao Labrador - vacinado em Almada, Lisboa
 Leia - cadela Cocker - vacinada em Condeixa, Coimbra
 Bounty - cadela - vacinada em Aveiro
 Tica - gata europeia - vacinada em Coimbra
 Tareco - gato - vacinado em Eiras, Coimbra

- Recorrendo a grupos, construa uma expressão regular que encontre os animais vacinados em **Coimbra**, em qualquer localidade. Indique os grupos necessários de forma a retirar o nome e o tipo de animal (apenas a indicação se é cão/cadela ou gato/gata).

73

73

Exercícios:

log01 - 12/01 - 09:00
 log02 - 05/01 - 06:10
 log02 - 05/01 - 06:66
 log03 - 07/01 - 01:99
 log99 - 19/01 - 08:57
 log123 - 21/01 - 09:01
 log1099 - 30/01 - 08:32
 log012 - 04/02 - 07:45

- Recorrendo a grupos, construa uma expressão regular que encontre os dias do mês de **Janeiro** em que os logs foram efetuados **antes das 9:00**. A ER regular deve encontrar apenas horas válidas, No ficheiro anterior há duas horas inválidas que não devem ser consideradas (06:66 e 01:99).

74

74