

A realizar: ☐ individualmente ☒ **em grupo**
Local de entrega: ☐ aula teórica ☒ **submissão electrónica**
Data limite entrega: **até às 15:30 do dia 16/Out**

OBJECTIVOS

Aprender a trabalhar com transdutores, usando-os para resolver um problema concreto na área de PLN.

ENUNCIADO

Javanaldo é um perigoso ladrão de livros vintage sobre a linguagem de programação Java. Rouba-os e contacta os cúmplices, através de emails codificados, indicando quando e onde será o próximo encontro. Estes recuperam os livros e vendem-nos, a preços brutais, a ávidos colecionadores de arte.

Ora, aquando do seu último furto, Javanaldo foi apanhado em flagrante delito pelo brilhante inspector Morcela. Lamentavelmente, a codificação sofisticada dos emails está a atrapalhar o trabalho do inspector, que necessita dos conhecimentos de um perito em Processamento de Língua Natural (PLN). Ora o maior especialista em PLN que Morcela conhece... és tu! Assim sendo, contacta-te e explica-te que cada email terá passado por três transdutores que, infelizmente, não foram encontrados. O primeiro transdutor converterá números Árábico decimal em números romanos; o segundo codificará os números romanos; o terceiro terá por alvo a codificação de letras. Felizmente, foi encontrada informação que permitirá reconstruir os transdutores. A tua tarefa é reconstruir os vários transdutores (individualmente) de modo a que seja possível:

1. Decodificar os dois últimos emails enviados por Javanaldo, indicando igualmente as versões intermédias destes emails (textos originais em anexo);
2. Codificar um novo email, escrito pelo inspector Morcela que será enviado aos cúmplices, na tentativa de os apanhar (igualmente em anexo). Mais uma vez deverão ser mostradas as versões intermédias.

Assim, tem de, pelo menos, desenvolver os seguintes transdutores:

Transdutor Romanos

Este transdutor converte um número Árábico decimal num número Romano (em maiúsculas, isto é, I, V, XVIII, etc.). Deverás fazer um conversor que funcione com todos os números de 1 até 100 e não pode aceitar mais nada que não sejam dígitos.

Exemplos:

"99" -> "XCIX"

"28" -> "XXVIII"

Transdutor 1

Este transdutor usa o transdutor anterior e processa frases com 1 ou mais "palavras", separadas pelo símbolo `_`. Cada palavra é uma sequência de dígitos ou letras minúsculas, não podendo ter simultaneamente letras e dígitos. Todas as frases terminam com o símbolo `_`. Neste transdutor os números arábicos são convertidos para números Romanos e mantém as outras palavras sem as alterar.

Exemplos:

"99_" -> "XCIX_"

"99_aa_" -> "XCIX_aa_"

"batata_28_" -> "batata_XXVIII_"

"ir_tambem_" -> "ir_tambem_"

Transdutor 2

Este transdutor processa frases e usa a seguinte chave de codificação para mascarar os números romanos, mantendo inalterados os restantes elementos:

I → 1 (I de Inês)

V → 2

X → 3

L → 4

C → 5

Exemplos:

"XCIX_" -> "3513_"

"XCIX_aa_" -> "3513_aa_"

"batata_XXVIII_" -> "batata_332111_"

"ir_tambem_" -> "ir_tambem_"

Transdutor 3

Este transdutor processa frases e usa a seguinte chave de codificação para converter algumas letras, mantendo inalterados os restantes elementos:

a → X

e → 9

i → V se seguido por um "r", L caso contrário

o → 0 (zero)

u → 8

mb → 7

Exemplos:

"3513_" -> "3513_"

"3513_aa_" -> "3513_XX_"

"batata_332111_" -> "bXtXtX_332111_"

"ir_tambem_" -> "Vr_tX79m_"

Transdutor codificador

Este transdutor aplica os três transdutores anteriores, pela ordem 1, 2, 3.

Exemplos:

"99_" -> "3513_"

"99_aa_" -> "3513_XX_"

"batata_28_" -> "bXtXtX_332111_"

"ir_tambem_" -> "Vr_tX79m_"

Transdutor descodificador

Este transdutor aplica os mesmos transdutores, mas pela ordem inversa: 3, 2, 1.

Exemplos:

"3513_" -> "99_"

"3513_XX_" -> "3513_aa_"

"bXtXtX_332111_" -> "batata_28_"

"Vr_tX79m_" -> "ir_tambem_"

Assume que:

- os emails descodificados contêm apenas minúsculas, dígitos e o símbolo _, que substitui os espaços em branco;
- todas as frases terminam em _;
- nunca ocorre um 0 (zero) sozinho;
- não serão considerados diacríticos ("'", "``", "~", "^", cedilha);
- o ficheiro "syms.txt" contém os símbolos a manipular pelos transdutores e não pode ser alterado.

SOFTWARE

Para testar a solução proposta utilize, em ambiente Linux, as ferramentas:

- "Graphviz" (<http://www.graphviz.org/>);
- "OpenFST" da Google (<http://www.openfst.org/twiki/bin/view/FST/FstDownload>).

Também são disponibilizados dois scripts, de uso opcional: *compact2fst.py* (gera um transdutor a partir de outro, escrito com uma notação mais compacta) e *word2fst.py* (gera um transdutor correspondente a uma palavra). Um ReadMe explica como usar estes scripts.

SUBMISSÃO

Submeta no Fenix, no projeto *MP1*, um ficheiro zip com:

- Os ficheiros de texto usados para definir os transdutores;
- O ficheiro run.sh com TODOS os comandos usados para gerar, quer em formato FST quer em formato gráfico (PS ou PDF): (i) os transdutores dos números romanos, 1, 2 e 3; (ii) o codificador; (iii) o decodificador; (iv) os dois emails a decodificar; e (v) o email a codificar;
- Um pequeno relatório (ficheiro txt ou pdf, MÁXIMO 1 página), contendo a identificação dos elementos do grupo, a descrição das opções tomadas e comentários à solução desenvolvida.

ATENÇÃO:

- O nome do ficheiro zip deve ter a forma numGrupo-X, em que X é "MP1-TAGUS" ou "MP1-ALAMEDA" (campus onde estão efetivamente a fazer a UC).
- Os transdutores desenvolvidos deverão ter os seguintes nomes:
 - transdutorRomanos.fst que APENAS transforma números em Árabe decimal (de 1 até 100) em números Romanos, falhando se lhe for passada qualquer letra;
 - transdutor1.fst, transdutor2.fst, transdutor3.fst consoante implementem as regras do transdutor 1, 2 ou 3;
 - codificador.fst e decodificador.fst. O primeiro consistirá em aplicar os três transdutores anteriores, pela ordem 1, 2, 3; o segundo pela ordem inversa.

ANEXO

Os dois últimos emails enviados por Javanaldo e que terás de decodificar são os seguintes:

Xs_21_pm_32111_d9_d9z97r0_3332111_3412_312_n_13_32111_31_0_
3332111_3412_321_n_13_2111_321_0_3311_d9_jXn9Vr0_p9IXs_312_h_

O novo email a codificar é:

biblioteca_do_palacio_de_monsserrate_na_estante_a_esquerda_no_dia_14_de_novembro_pelas_10_23_de_maio_

CRITÉRIOS DE AVALIAÇÃO

Na avaliação serão tidos em conta os seguintes critérios (máximo = 4 valores):

1. Utilização de vários transdutores, evitando repetições desnecessárias (0,4 valores);
2. Correcto funcionamento dos transdutores: transdutorRomanos.fst, transdutor1.fst, transdutor2.fst e transdutor3.fst (0,4 valores cada);
3. Correcto funcionamento dos transdutores: codificador.fst e decodificador.fst (0,5 valores cada);
4. Correcto funcionamento do "run.sh" (0,4 valores);
5. Entrega das versões gráficas de todos os transdutores, bem como dos emails a decodificar e do emails a codificar, nas suas diferentes formas, isto é, após passarem por cada um dos transdutores (0,2 valores);
6. Qualidade do relatório (0,2 valores);
7. Correção ortográfica e sintáctica do relatório (0,2 valores);

O não cumprimento de qualquer regra implica um desconto mínimo de 2 valores (em 4 valores).

Both instructors and students can consider steps to enhance academic integrity in the CMU community. This section offers suggestions drawn from ongoing conversations with CMU students and faculty over the years and from the literature on academic integrity. The steps below include ways students can more effectively manage their own learning with the help of university resources and ways individual instructors can enhance support for student learning and integrity.

Steps Students Might Take:

- Ask about policies regarding collaboration and citations at the beginning of each course. Instructors' policies may differ substantially from one another.
- Ask questions - in class, immediately after class, in e-mail or in office hours - about course content or course procedures. If you are confused, you might ask for more clarification, different examples, or specific applications to help you understand. Other students often have the same questions you do so your questions can enhance the overall effectiveness of the course.
- Find out whether the instructor will provide suggestions for preparing for exams and consider preparing your own review sheet. The process of making a review sheet is actually a good method of improving your understanding of and memory for complex information.
- Refine your note-taking skills. Many students form the habit of transcribing whatever the professor writes, no more and no less. To facilitate better review and study sessions, ask yourself frequent questions as you read or listen to a lecture: What is the key new idea here? How can I use this information? Can I anticipate what is coming next?
- Improve your time management, especially during the day and early evening. Procrastination more often leads to ineffective cramming and loss of sleep than to good performance under pressure. If you begin to work well before due dates and examinations, you are much more likely to learn the material, to be able to get help if you need it, to feel less stressed, to perform better, and to avoid poor decisions on very late nights.
- Speak with the professors about their grading and homework policies if you feel that the policies seem unfair - feedback is essential to improving the quality of a class. If you feel uncomfortable talking with an instructor directly, you might express your views in early course evaluations or to a teaching assistant.