



Nome: _____ Turma: _____

Data: ____ / ____ / ____

Nota

Questão 1 (5 pontos)

M-Palavras maiúsculas; m-Palavras minúsculas; N-números; <separadores, idênticos ao original>; p-Preposições; c-Conjunções; a-Artigos; O-Ordinais (números seguidos pelos caracteres ^o e ^a).

Exemplo de texto (apenas para conferência), tokens de entrada e string de saída codificada:

"Amanha, Companhia Vale do Rio Doce roda d'agua: 'e assim foi' 1º V.Sra. V.Exma."

(iii) 'M, m M p M M m m'm: 'm m m' O M.M. M.M.'

O tokenizador de nível inclui o caracter espaço como token. Sequencias de mais de 1 espaço seguidos são considerados como apenas 1 espaço.

Construa a função `carregaLst(<nome do arquivo>,)` que recebe como parâmetro de entrada o nome de um arquivo texto contendo palavras, 1 palavra em cada linha. A função deve retornar uma lista contendo as palavras encontradas no arquivo. Tratamento de exceção deve ser feito para verificar a existência prévia do arquivo e evitar erros de execução.

Questão 3 (10 pontos)

Construa a função *tokenizadorN2* (<tokens>, <codificação>, <pronomes tratamento>) que recebe como parâmetros de entradas uma lista de tokens, sua respectiva string codificada (saída da função da questão 1) e uma lista de pronomes de tratamento. A função *tokenizadorN2* deve produzir como saída uma lista de tokens com 1 nível de abstração acima dos tokens de nível 1. Nesse novo nível de abstração são tratados como tokens as seguintes entidades de texto: pronomes de tratamento, datas, horas, palavras aglutinadas com aspas (como rod'agua e sat'anna) e citações. É obrigatório que a função deve fazer o seu trabalho atuando nos tokens de nível 1 a partir da string codificada correspondente, como analisado e desenvolvido nas aulas de laboratório.

Exemplo do texto original, tokens de entrada (nível 1), string codificada e tokens de nível 2, saída da função pedida:

(i) Texto original (aspas duplas não incluídas e apenas para conferência):

"Amanha, Companhia Vale do Rio Doce roda d'agua: 'e assim foi' 1º V.Sra. V.Exma. 27/05/2015."

(ii) ['Amanha', ',', ',', 'Companhia', ',', 'Vale', ',', 'do', ',', 'Rio', ',', 'Doce', ',', 'roda', ',', 'd', '','', 'agua', ':', ',', '','', 'e', ',', 'assim', ',', 'foi', '','', '1º', ',', 'V', ',', 'Sra', ',', ',', 'V', ',', 'Exma', ',', ',', '27', '/', '05', '/', '2015']

(iii) 'M, m M p M M m m:m: 'm m m' O M.M. M.M. N/N/N.'

(iii) ['Amanha', ',', ',', 'Companhia', ',', 'Vale', ',', 'do', ',', 'Rio', ',', 'Doce', ',', 'roda', ',', '“d"agua”, ':', ',', '','', 'e assim foi', ',', '1º', ',', 'V.Sra.', ',', 'V.Exma.', ',', '27/05/2015']

Questão 4 (5 pontos)

Construa uma aplicação que utilize as funções desenvolvidas nas questões anteriores para executar as seguintes tarefas: ler as bases de dados de conjunções, preposições, pronomes de tratamento e artigos e gerar suas respectivas listas (utilize as bases de dados fornecidas com a prova: arquivos *preposicoes.txt*, *artigos.txt*, *conjuncoes.txt* e *prontratamento.txt*); ler uma string de texto fornecida pelo usuário; gerar a tokenização de nível 2; imprimir na saída o texto original de entrada e a tokenização de nível 2.

Importante: utilize apenas os recursos de python vistos em aula com o professor. A questão é considerada correta se produzir o resultado correto e for construída segundo a especificação constante no enunciado. Recursos de que não foram vistos em aula, como expressões regulares e rotinas condensadas de strings, não devem ser utilizados na construção das soluções. Utilize apenas o que foi visto em aula e seus conhecimentos de lógica.

Na dúvida, consulte o professor.

Boa Prova.