

Trabalho 2 – Funções, Vetores e Matrizes

Introdução a Ciência da Computação 1

Prof. Rudinei Goularte e Moacir Ponti

Monitores/PAE: Tamires T.S. Barbieri, Henrico Brum, Lucas T. Da Silva, André Luis M. Fakhoury e Gabriel Van Loon

Descrição do Problema:

Você conseguiu um maravilhoso estágio no site de e-commerce Buscapé. Infelizmente, todos na empresa desprezaram suas excelentes qualidades de profissional e te colocaram na posição de identificador de novas tendências e estratégias de *marketing* digital na empresa. Sua principal atribuição é ler comentários na internet sobre produtos eletrônicos e participar de reuniões respondendo questões como ‘quantas pessoas gostaram do novo celular da Apple?’ ou ‘quantas vezes alguém usou a palavra *ruim* em um comentário sobre o televisor LG’.

O trabalho vai muito bem, diariamente você recebe alguns comentários para ler e algumas perguntas para responder na reunião do próximo dia. Entretanto, você é um exímio programador computacional e, além disso, seu colega de faculdade lhe indicou a série *Lost* e agora você está completamente viciado... portanto você teve uma ideia sensacional.

Por que em vez de ler cada comentário manualmente você não cria um programa em C que receba todos os comentários diários de cada produto e os processa para te mostrar, com um simples comando, as informações pedidas na reunião do próximo dia? Que ideia genial! Agora você vai poder descobrir os segredos da ilha e ainda vai passar uma imagem de profissionalismo na reunião matinal!

Claro que para isso você precisa construir esse programa e para isso você procura na internet maneiras de te ajudar com isso. A sorte é que as perguntas geralmente são as mesmas, então você sabe o que podem te perguntar e como transformar isso em código.

Seu programa receberá como entrada diversas sentenças em português brasileiro, já processadas e sem pontuação. O número de sentenças e o número de perguntas é o cabeçalho da entrada. Confira o exemplo abaixo.

3 2 Televisor LG ; adorei o televisor lg ele e muito funcional Ø Laptop HP ; nao gostei desse notebook pois ele esquentava muito Ø Celular Motorola ; o formato do celular e muito bom Ø palavra bom em Celular Motorola quantos positivos em Televisor LG	<n. de sentenças> <n. de perguntas> <nome_do_produto> ; <sentença> <nome_do_produto> ; <sentença> <nome_do_produto> ; <sentença> <pergunta 1> <pergunta n+1>
---	---

Toda a entrada terá o tipo do produto e a marca, um demarcador e um comentário com no **máximo** 20 palavras (cada palavra com no máximo 19 caracteres). Haverá um identificador de fim de sentença no final de cada comentário ('Ø'). O pessoal do pré-processamento também removeu todas as pontuações, então fique tranquilo!

Durante a reunião, geralmente os supervisores só perguntam duas coisas – segue abaixo a definição de ambas:

1. Estagiário, quantas vezes os usuários usam a palavra 'perfeito' em um comentário sobre o 'Celular Motorola'?

Para responder essa pergunta você descobre que existe uma técnica muito simples que pode ser aplicada chamada *Bag-of-words* onde você vai representar cada sentença como um **vetor** de contagem de palavras presentes em um vocabulário, por exemplo:

Sentenças	Representação Bag-of-words								
	gostei	desse	celular	não	laptop	adorei	esse	monitor	e
gostei desse celular	1	1	1	0	0	0	0	0	0
não gostei desse laptop	1	1	0	1	1	0	0	0	0
adorei esse monitor e esse laptop	0	0	0	0	1	1	2	1	1

Usando essa abordagem, você pode fazer um vetor *Bag-of-words* para cada **conjunto de comentários** sobre **cada produto** e você só precisa indicar uma palavra e um tipo de produto para receber como saída a ocorrência dessa palavra. Nesse trabalho haverão no máximo 200 palavras diferentes ao todo, ou seja, tamanho máximo do BoW=200.

Exemplo:

Input: 3 1 Televisor LG ; adorei o televisor lg ele e muito funcional Ø Laptop HP ; nao gostei desse notebook pois ele esquenta muito Ø Celular Motorola ; o formato do celular e muito bom bom mesmo bom demais Ø palavra bom em Celular Motorola	Output: 3
--	---------------------

Porém, outra pergunta geralmente é feita... mas essa também é bem fácil de responder...

2. Querido estagiário, quantas pessoas estão falando bem do 'Laptop HP'?

Para responder essa pergunta você tem um trunfo na manga!

De tanto ler comentários você identificou que as pessoas sempre elogiam os produtos com as palavras:

[bom, maravilhoso, otimo, sensacional, excelente, adorei, gostei, amei, eficiente, boa, maravilhosa, otima]

Enquanto que você só encontra usuários insatisfeitos usando os termos:

[detestei, odiei, ruim, pessimo, terrivel, raiva, odio, pessima, lento, lenta, fragil, desisti]

Nesse caso você poderia só identificar os reviews que têm esses termos. observe o seguinte exemplo:

Input: 4 1 Impressora Epson ; funcionar nao funciona depois de um tempo simplesmente desisti Ø Impressora Epson ; a impressora nao funciona de jeito nenhum odiei Ø Televisor LG ; o televisor e excelente gostei muito da definição da tela Ø Impressora Epson ; adorei a versao de jato de tinta nao tenho reclamacoes Ø quantos negativos em Impressora Epson	Output: 66.7%
---	-------------------------

Obviamente você recebe diariamente muito mais que 4 sentenças, assim como muito mais do que uma pergunta. Apesar disso, você recebe sempre o mesmo número de produtos (10). O programa deve ser capaz de buscar **qualquer termo** nas sentenças, assim como identificar a porcentagem de sentenças **Positivas ou Negativas**. Lembre-se que a consulta muda dependendo do nome do produto – **respeite as entradas!**

[Bônus] – Invertendo Opiniões

Depois de muito tempo usando esse programa você notou que em alguns casos os usuários usavam advérbios de negação nos comentários. Como você não sabe nada de português, ignorou por um bom tempo... porém isso começou a complicar a sua classificação de polaridades! Os advérbios de negação modificam a sentença **invertendo a polaridade**. Os advérbios são:

[nao, jamais, nunca]

O maior problema é que esses advérbios podem aparecer em qualquer posição da sentença, mas eles só invertem a polaridade se estiverem até 3 posições **antes** da palavra que representa polaridade. Por exemplo:

Input: 4 1 Impressora Epson ; nao comprem ela e um nojo <u>jamais gostei</u> dela Ø Laptop HP ; que latop lento meu deus do ceu Ø Impressora Epson ; a impressora <u>nao</u> e muito <u>boa</u> Ø Impressora Epson ; nem sei a performance <u>nunca</u> pareceu <u>eficiente</u> Ø quantos negativos em Impressora Epson	Output: 100.0%
--	--------------------------

Instruções

O código fonte (apenas arquivo .c) deve ser submetido pelo <http://run.codes>.

IMPORTANTE: os dois últimos casos de teste são referentes à questão bônus. Se você quiser implementar sem a questão bônus, ignore o resultado dos **dois últimos casos de teste**.

O trabalho será avaliado levando em consideração:

1. realização dos objetivos;
2. representação correta da entrada e saída de dados;
3. uso de comentários e estrutura do código (ex.: indentação, legibilidade);
4. número de acertos no sistema run.codes nos primeiros oito casos de teste;
5. um bônus será concedido a quem conseguir resolver todos os casos de teste.