

MC102 - Algoritmos e Programação de Computadores**Turmas QRSTWY****Instituto de Computação - Unicamp****Professores:** Hélio Pedrini e Zanoni Dias**Monitores:** Andre Rodrigues Oliveira, Gustavo Rodrigues Galvão, Javier Alvaro Vargas Muñoz e Thierry Pinheiro Moreira

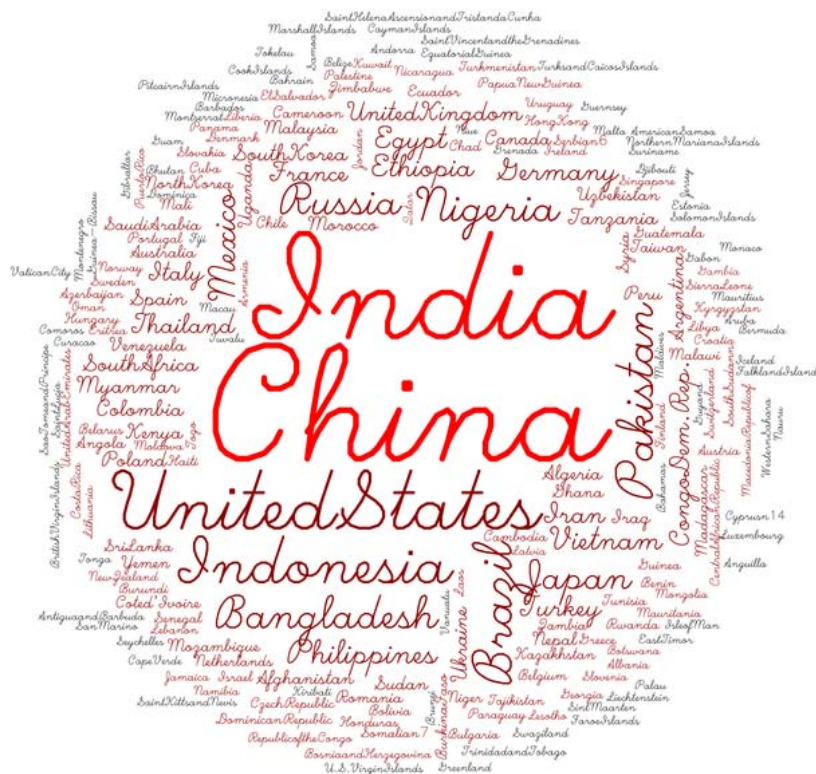
Lab 10a - Tag Cloud

Prazo de entrega: 25/05/2015 às 13h59m59s**Peso:** 7

Tag cloud é uma representação visual dos dados de um texto, utilizada de forma geral para descrever palavras-chaves em websites. As tags são tipicamente palavras, e a importância de cada tag é visualizada através de seu tamanho ou de sua cor. Este formato é útil para perceber os termos mais proeminentes e sua importância relativa em um texto.

Seu visual, sua adoção e o uso do termo "tag cloud" ganhou destaque na primeira década do século 21, como uma característica dos primeiros sites e blogs da Web 2.0, e eram utilizados primeiramente para a visualização da distribuição dos metadados que descreviam o conteúdo do site.

Mais tarde, a tag cloud foi ganhando outros usos, como gerar estatísticas de textos e até mesmo representação visual de estatísticas populacionais e geográficas, como mostra a figura abaixo a distribuição da população mundial de 2007, onde o tamanho da fonte de um país mostra o tamanho de sua população com relação aos demais países.



O Flickr foi o primeiro grande site a adotar a tag cloud, em 2004, ajudando inclusive em sua divulgação e adoção em massa por outros sites. Entretanto, a saturação de seu uso e as dúvidas com relação a sua utilidade como ferramenta de navegação web levaram a uma queda notável de seu uso dois anos mais tarde. Depois desta queda, as tag clouds passaram a ser utilizadas para visualização básica de estatísticas e análise de textos.

O conteúdo das tag clouds é único: palavras. O que muda é a aparência: sua organização, a fonte utilizada e as cores. Geralmente, além de maiores, as tags mais frequentes de um texto costumam ter cores mais fortes, como azul e preto, enquanto que tags menos frequentes e, por este motivo menores, costumam ter cores mais acizentadas. As tag clouds ajudam a detectar, visualmente, as diferenças e similaridades entre dois textos, como podemos ver na imagem abaixo, que compara os discursos *State of the Union* apresentado por George W. Bush em 2002 e o apresentado por Barack Obama em 2011. As palavras "american", "jobs", "people", "nation" e "work" possuem frequências similares nos discursos dos dois presidentes, enquanto "freedom", "terror", "terrorists", "war" e "weapon" são muito mais frequentes no discurso do Bush, de 2002.



Para gerar uma tag cloud, deve-se primeiro contar as ocorrências de todas as palavras do texto, a fim de saber quais são as palavras que ganharão destaque na representação visual. Sua tarefa neste laboratório será, a partir de um texto (formado apenas por palavras), retornar as palavras em ordem decrescente de número de ocorrências.

Entrada

- A primeira linha da entrada consistirá de 4 inteiros, D F L M , onde:
 - D é o número máximo de palavras distintas no texto;
 - F é o número mínimo de ocorrências para uma palavra ser considerada nas estatísticas;
 - L é o tamanho mínimo de uma palavra para que ela seja considerada nas estatísticas.
 - M é o tamanho máximo de uma palavra do texto.
- As demais linhas consistem no texto a ser processado, sem letras maiúsculas, sem acentuação e sem pontuação. O fim do texto será indicado pelo caractere '*'.

Saída

- Seu programa deve imprimir a lista de palavras com pelo menos L caracteres e F ocorrências, uma por linha, no formato ' $w_i \ q_i$ ', onde q_i é o número de ocorrências da palavra w_i no texto, em

ordem decrescente de número de ocorrências q.

- o Caso duas ou mais palavras possuam o mesmo número de ocorrências q, seu programa deve imprimi-las em ordem lexicográfica.

Exemplos

#	Entrada	Saída
1	<p>80 2 4 12</p> <p>graduacao</p> <p>e um grau do ensino superior obtido em cursos de graduacao ministrados pela universidade destinados a alunos que tenham concluido o ensino medio ou equivalente o ingresso se da por processo seletivo</p> <p>a unicamp oferece cursos de graduacao em diferentes areas do conhecimento biologicas e profissoes de saude artes humanas exatas tecnologicas e da terra</p> <p>a estrutura de ensino e pesquisa da unicamp esta em sua maior parte no campus de barao geraldo em campinas em piracicaba funciona a faculdade de odontologia de piracicaba fop em limeira a faculdade de tecnologia ft e faculdade de ciencias aplicadas fca em paulinia o centro de pesquisas quimicas biologicas e agricolas cpqba em sumare o hospital estadual e em hortolandia o hospital mario covas</p> <p>*</p>	<p>ensino 3</p> <p>faculdade 3</p> <p>graduacao 3</p> <p>biologicas 2</p> <p>cursos 2</p> <p>hospital 2</p> <p>piracicaba 2</p> <p>unicamp 2</p>
2	<p>80 2 4 14</p> <p>opcoes curriculares</p> <p>na matricula inicial na universidade e atribuido ao aluno um curriculo pleno do ano de seu ingresso</p> <p>em curso no qual haja mais de uma habilitacao ou enfase a coordenadoria de graduacao estabelece aquela em que o aluno sera matriculado inicialmente</p> <p>para cursos nos quais a opcao por lingua estrangeira e um componente curricular obrigatorio o aluno deve fazer a sua opcao pela lingua no ato da matricula na universidade</p> <p>no decorrer do curso o aluno pode alterar o curriculo pleno a habilitacao ou enfase e a lingua estrangeira para determinadas habilitacoes ou enfases ha a necessidade de aprovacao da coordenadoria de graduacao</p> <p>a opcao por curriculo pleno anterior ao que o aluno vem cumprindo depende em qualquer caso da aprovacao das coordenadorias de graduacao responsaveis pelas disciplinas faltantes para a conclusao de seu curso pelo curriculo pleno solicitado</p> <p>*</p>	<p>aluno 5</p> <p>curriculo 4</p> <p>pleno 4</p> <p>curso 3</p> <p>graduacao 3</p> <p>lingua 3</p> <p>para 3</p> <p>aprovacao 2</p> <p>coordenadoria 2</p> <p>enfase 2</p> <p>estrangeira 2</p> <p>habilitacao 2</p> <p>matricula 2</p> <p>opcao 2</p> <p>universidade 2</p>
3	<p>120 3 3 16</p> <p>artigo 1 os cursos de graduacao tem por finalidade habilitar a obtencao de graus academicos ou graus que correspondem a profissoes regulamentadas em lei</p> <p>paragrafo unico todos os cursos de graduacao da unicamp serao regidos por esta deliberacao</p> <p>artigo 2 cada curso de graduacao e coordenado e supervisionado em suas atividades de ensino por uma comissao de graduacao presidida por um professor doutor denominado coordenador de curso</p>	<p>cursos 6</p> <p>graduacao 6</p> <p>por 6</p> <p>artigo 4</p> <p>atividades 3</p> <p>coordenador 3</p> <p>curso 3</p> <p>unicamp 3</p>

	<p>paragrafo unico o coordenador de curso sera auxiliado por um professor doutor denominado coordenador associado com quem compartilhara as atividades de coordenacao e que substituirá em seus impedimentos</p> <p>artigo 3 os cursos de graduacao poderao ser ministrados em turno integral matutino vespertino e noturno</p> <p>artigo 4 o sistema de integralizacao dos cursos adotado pela unicamp e de creditos e de matriculas por disciplina</p> <p>1 o credito e a unidade de medida do trabalho escolar dos cursos de graduacao da unicamp e corresponde a 15 quinze horas aula de atividades academicas</p> <p>2 as unidades de ensino deverao estabelecer limites maximos de creditos para efeito de matricula nos cursos ministrados sob sua responsabilidade</p> <p>*</p>	
4	<p>52 1 10 14</p> <p>integralizacao</p> <p>integralizacao de um curso e o cumprimento de todas as exigencias fixadas no catalogo de cursos de graduacao</p> <p>o sistema de integralizacao adotado pela universidade e o de credito e matricula por disciplina</p> <p>para informacao do aluno e publicado na web semestralmente o relatorio de integralizacao curricular no qual constam dentre outras informacoes as disciplinas cursadas com seus respectivos creditos e o prazo maximo para integralizacao</p> <p>o aluno pode consultar em servicos online seu relatorio de integralizacao curricular</p> <p>*</p>	<p>integralizacao 6</p> <p>curricular 2</p> <p>cumprimento 1</p> <p>disciplina 1</p> <p>disciplinas 1</p> <p>exigencias 1</p> <p>informacao 1</p> <p>informacoes 1</p> <p>respectivos 1</p> <p>semestralmente 1</p> <p>universidade 1</p>
5	<p>72 2 3 14</p> <p>instituto de computacao</p> <p>informacoes sobre mestrado e doutorado</p> <p>o programa de pos graduacao do ic teve inicio com o mestrado em ciencia da computacao em 1977</p> <p>o doutorado em ciencia da computacao iniciou se em 1993 ambos os cursos possuem nivel 7 na capes e ja formaram mais de 700 mestres e 150 doutores</p> <p>o programa prima pela formacao e producao de teses e dissertacoes com alta qualidade o que pode ser atestado pela quantidade de artigos com co autoria de alunos do programa e premios recebidos em concursos de teses e dissertacoes e de melhores trabalhos em congressos nacionais e internacionais</p> <p>foram mais de 85 premios obtidos desde 1997 pelos alunos e 42 pelos professores</p> <p>*</p>	<p>com 3</p> <p>computacao 3</p> <p>programa 3</p> <p>alunos 2</p> <p>ciencia 2</p> <p>dissertacoes 2</p> <p>doutorado 2</p> <p>mais 2</p> <p>mestrado 2</p> <p>pela 2</p> <p>pelos 2</p> <p>premios 2</p> <p>teses 2</p>