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

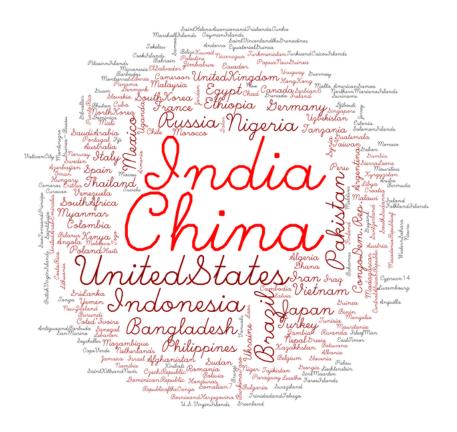
Prazo de entrega: 25/05/2015 às 13h59m59s

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.

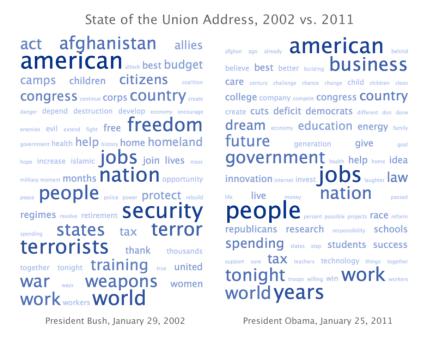
Seu visual, sua adoção e o uso do termo "tag cloud" ganhou destaque na primeira década do século 21, como uma característica dos primeiros sites e blogs da Web 2.0, e eram utilizados primeiramente para a visualização da distribuição dos metadados que descreviam o conteúdo do site.

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 frequêntes 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, afim 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.

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

Exemplos

#	Entrada	Saída
1	80 2 4 12 graduacao 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 a unicamp oferece cursos de graduacao em diferentes areas do conhecimento biologicas e profissoes de saude artes humanas exatas tecnologicas e da terra 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 *	ensino 3 faculdade 3 graduacao 3 biologicas 2 cursos 2 hospital 2 piracicaba 2 unicamp 2
2	80 2 4 14 opcoes curriculares na matricula inicial na universidade e atribuido ao aluno um curriculo pleno do ano de seu ingresso em curso no qual haja mais de uma habilitacao ou enfase a coordenadoria de graduacao estabelece aquela em que o aluno sera matriculado inicialmente para cursos nos quais a opcao por lingua estrangeira e um componente curricular obrigatorio o aluno deve fazer a sua opaoo pela lingua no ato da matricula na universidade 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 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 *	aluno 5 curriculo 4 pleno 4 curso 3 graduacao 3 lingua 3 para 3 aprovacao 2 coordenadoria 2 enfase 2 estrangeira 2 habilitacao 2 matricula 2 opcao 2 universidade 2
3	120 3 3 16 artigo 1 os cursos de graduacao tem por finalidade habilitar a obtencao de graus academicos ou graus que correspondem a profissoes regulamentadas em lei paragrafo unico todos os cursos de graduacao da unicamp serao regidos por esta deliberacao artigo 2 cada curso de graduacao e coordenado e supervisionado em suas atividades de ensino por uma comissao	cursos 6 graduacao 6 por 6 artigo 4 atividades 3 coordenador 3 curso 3 unicamp 3

de graduacao presidida por um professor doutor denomidado coordenador de curso

paragrafo unico o coordenador de curso sera auxiliado por um professor doutor denominado coordenador associado com quem compartilhara as atividades de coordenacao e que substituira em seus impedimentos

artigo 3 os cursos de graduacao poderao ser ministrados em turno integral matutino vespertino e noturno artigo 4 o sistema de integralizacao dos cursos adotado pela unicamp e de creditos e de matriculas por disciplina 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

2 as unidades de ensino deverao estabelecer limites maximos de creditos para efeito de matricula nos cursos ministrados sob sua responsabilidade

*

4 52 1 10 14

integralizacao

integralizacao de um curso e o cumprimento de todas as exigencias fixadas no catalogo de cursos de graduacao o sistema de integralizacao adotado pela universidade e o de credito e matricula por disciplina 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

o aluno pode consultar em servicos online seu relatorio de

integralizacao 6
curricular 2
cumprimento 1
disciplina 1
disciplinas 1
exigencias 1
informacao 1
informacoes 1
respectivos 1
semestralmente 1
universidade 1

5 72 2 3 14

instituto de computacao

integralizacao curricular

informacoes sobre mestrado e doutorado

- o programa de pos graduacao do ic teve inicio com o mestrado em ciencia da computacao em 1977
- 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
- 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 foram mais de 85 premios obtidos desde 1997 pelos alunos e 42 pelos professores

com 3
computacao 3
programa 3
alunos 2
ciencia 2
dissertacoes 2
doutorado 2
mais 2
mestrado 2
pela 2
pelos 2
premios 2
teses 2