

Universidade de Brasília  
Departamento de Ciência da Computação  
Projeto 5 , Estruturas de Dados, Turma E, 2015/1  
Prof. Dúbio

*João amava Teresa que amava Raimundo  
que amava Maria que amava Joaquim que amava Lili  
que não amava ninguém.  
João foi para os Estados Unidos, Teresa para o convento,  
Raimundo morreu de desastre, Maria ficou para tia,  
Joaquim suicidou-se e Lili casou com J. Pinto Fernandes  
que não tinha entrado na história.*

(ANDRADE, CARLOS DRUMMOND DE. “Quadrilha”. In *Alguma Poesia*, Edições Pindorama, Rio de Janeiro, 1930)

Grafos são estruturas de dados com muitas aplicações e possibilidades de representação de inúmeros processos na vida real. A rede internet pode ser representada como um grafo, onde informações e programas são dispostos em portais, páginas e acessados por computadores em todo o planeta. Um dos sucessos de ferramentas na rede internet é chamado de “máquina de busca”, um programa que disponibilizado para todos os nós da rede recebe um texto e busca e retorna ao usuário uma sequência de endereços (nós), indicando quais na ordem apresentada podem ter maior relevância ao texto dado como entrada. Uma das empresas de maior sucesso financeiro atualmente surgiu construindo um programa desse tipo, a “google” (<http://infolab.stanford.edu/~backrub/google.html>).

O algoritmo de ranqueamento das páginas web utiliza como princípio de popularidade de uma página (A) o número de páginas (i.e. URLs) que direcionam visitas a esta, como na fórmula (1)

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn)) \quad (1)$$

onde PR (*Page Rank*), T... (T páginas que apontam para página A), C (T número de páginas que são direcionadas pela página T), d (fator de ajuste entre 0 e 1.0, sendo usado heurísticamente como 0.85) (ps. Mais detalhes fornecidos nas notas de aula)

Neste projeto pede-se escrever um programa em C que considere duas entradas. A primeira será um arquivo fornecido *allpages.txt*, que contém todos os endereços *html* (páginas) que devem ser pesquisadas para elaborar um “*ranking*”. Um outro arquivo contendo um conjunto de páginas de uma determinada universidade, *sunysb.tar.gz*, deve ser aberto e colocado em um diretório para que as páginas fornecidas em *allpages.txt* sejam pesquisadas nesse diretório e um mapa de influência com os valores calculados pelo PR (Page Rank) seja fornecido. O número de nós deve ser o número de páginas fornecidas em *allpages.txt*, de uma estrutura de grafo orientado, implementado considerando lista de adjacências onde cada nó possui um rótulo do tipo *char\** (string) e um posto (ranking) do tipo *float*, e seus adjacentes. Após o grafo (“mini-web”) ter sido lido o programa deve calcular, usando *n* rodadas, os postos (ou *page ranks*) de cada nó e no final imprimir em um arquivo (“*rankingED20151.txt*”) e na tela do computador as listas de adjacências dos nós do grafo, e os postos finais de cada, em ordem decrescente.

O código deve ser bem documentado, de forma modular com funções para cada tarefa independente, realizado por dois (2) estudantes do curso usando *“pair programming”*, e entregue via sistema <http://aprender.unb.br> do curso, no prazo estipulado.