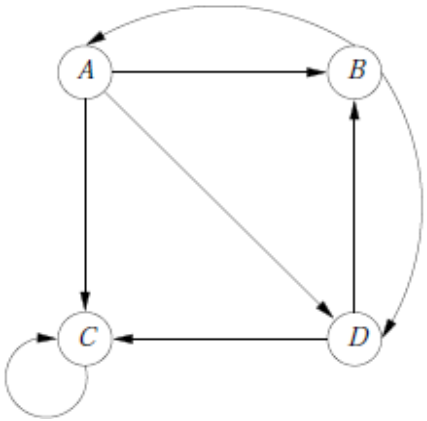
Programming Assignment # 1

PageRank

$$PR(A) = \frac{(1-d)}{N} + d * (\frac{PR(T_1)}{C(T_1)} + ... + \frac{PR(T_n)}{C(T_n)})$$



$$\mathbf{v}' = \begin{bmatrix} 0 & 2/5 & 0 & 0 \\ 4/15 & 0 & 0 & 2/5 \\ 4/15 & 0 & 4/5 & 2/5 \\ 4/15 & 2/5 & 0 & 0 \end{bmatrix} \mathbf{v} + \begin{bmatrix} 1/20 \\ 1/20 \\ 1/20 \\ 1/20 \end{bmatrix}$$

$$\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}, \begin{bmatrix} 9/60 \\ 13/60 \\ 25/60 \\ 13/60 \end{bmatrix}, \begin{bmatrix} 41/300 \\ 53/300 \\ 153/300 \\ 53/300 \end{bmatrix}, \begin{bmatrix} 543/4500 \\ 707/4500 \\ 2543/4500 \\ 707/4500 \end{bmatrix}, \ldots, \begin{bmatrix} 15/148 \\ 19/148 \\ 95/148 \\ 19/148 \end{bmatrix}$$

For A, the new page rank equals
$(0, \frac{2}{5}, 0, 0)(PR(A), PR(B), PR(C), PR(D))^T + 0.2 \times \frac{1}{4}$

Computing PageRank

PageRank(page(0..$n-1$), d, $DIFF$);

1 INITIALIZE:
$PR(i) = 1/N$ for all i;

2 diff = 0;
For each P in Page
PRbefore= $PR(P)$;
$PR(P) = (1-d)/N + d * ((PR(t_1)/CR(t_1) + PR(t_2)/CR(t_2)...)$
for all t_i that point to P
($CR(t_i)$ is the outbranching of t_i)
$diff = diff + AbsoluteValue(PRbefore - PR(P)$;

3 if $diff >= DIFF$ then goto 2

4 done.

Outputs

- Output #1: Page Rank list
  A file with a list of the pages in page rank order (highest to lowest), for each page listing its outbranching (how many pages it connects to) and the page rank for that page (to 8 significant digits).
  Example

page500	137	.0023781
page503	196	.0022139
page1500	7	.0021101

Outputs

- Output #2: Reverse index
  A file which shows its reverse index. That is, for every word you should output the word followed by a list of the pages that include that word (please alphabetize the word list, the pages can be produced in any order you choose - but each page should only appear once!).

  Example

  a page500 page522 page916 page803
  an page999 page921 page998 page763 page1227 page501 page93
  page522 page916 page803 page919 page997
  at page900 page2000 page1227 page501 page937 page862
  zuchi page1000

Outputs

- Output #3: Search engine
  Take as input a list of words and you output the top ten pages (sorted by page rank) that contain them (if less than 10, then print all of them). When a single word is entered, just output the pages for that word. When multiple words are entered, output two lists - the top ten pages with **all** the words and the top ten pages with **any** of the words (multiple words do not effect the ranking here).
  Example
  1. Enter Word: moose
     page1000 page2001 page2000
  2. Enter Word: moose cow
     AND (moose cow) page1000
     OR (moose cow) page1000 page2006 page2001 page1983 page776 page842 page777 page2000 page963 page871

An Example

page351 page352 page119 page354 page449 page445 page111 page440 page117 page114 page115 page219 page191 page196 page197 page195 page213 page215 page214 page290 page292 page65 page297 page299 page298 page207 page447 page472 page475 page362 page163 page160 page147 page58 page288 page286 page287 page284 page51 page273 page378 page275 page373 page279 page278 page178 page177 page172 page48 page47 page46 page44 page312 page260 page454 page265 page266 page268 page269 page149 page148 page140 page491 page497 page144 page301 page494 page383 page381 page466 page69 page118 page460 page36 page35 page33 page31 page90 page154 page94 page95 page97 page98 page257 page256 page158 page489 page319 page306 page480 page316 page484 page143 page486 page397 page478 page331 page399 page411 page28 page415 page25 page27 page227 page416 page123 page87 page86 page241 page83 page328 page321 page320 ——————— her clearing instrument of singular every lenses and He the him at which false successfully to the and In Street