

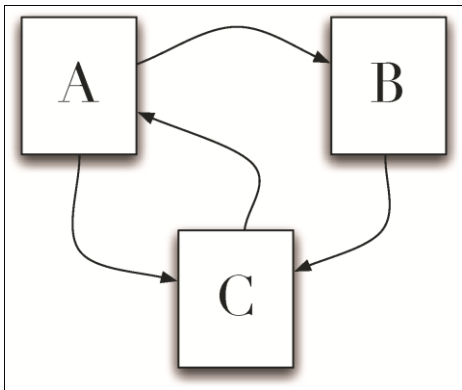
## Search Engine Google

- menyediakan layanan untuk mencari informasi pada database berisi milyaran halaman web.
- Google membantu mensortir halaman – halaman yang diminta sesuai dengan ‘rank’ tiap halaman.
- Google mampu menandai tiap halaman dengan sebuah rank, lalu menampilkan halaman hasil pencarian sesuai dengan rank tertinggi sampai terendah.
- Tahap – tahap proses pencarian pada Google :
  1. Membaca dan memilah query.
  2. Mengubah kalimat menjadi wordID.
  3. Mencari kata dokumen yang dicari berdasarkan query dalam database secara singkat.
  4. Mencari dari hasil proses ketiga ,apakah ada yang mengandung semua kriteria secara utuh.
  5. Menghitung rank dari dokumen tersebut.
  6. Jika dokumen tidak ditemukan, kembali ke proses keempat dengan mencari dalam database secara penuh ( full scan ).
  7. Jika belum selesai ,kembali ke step 4.
  8. Mengurutkan dokumen berdasarkan rank-nya lalu selesaikan proses.

### Algoritma PageRank

- PageRank ditentukan untuk setiap halaman bukan keseluruhan situs web.
- PageRank sebuah halaman ditentukan dari PageRank halaman yang mengacu kepadanya
- Jadi proses ini akan berulang sampai ditemukan hasil yang tepat.
- PageRank halaman A tidak langsung diberikan kepada halaman yang dituju, tetapi sebelumnya dibagi dengan jumlah link yang ada pada halaman T1 (outbound link), dan PageRank itu akan dibagi rata kepada setiap link yang ada pada halaman tersebut.
- Demikian juga dengan setiap halaman lain “Tn” yang mengacu ke halaman “A”. Setelah semua PageRank yang didapat dari halaman-halaman lain yang mengacu ke halaman “A” dijumlahkan, nilai itu kemudian dikalikan dengan damping faktor yang bernilai antara 0 sampai 1.
- Hal ini dilakukan agar tidak keseluruhan nilai PageRank halaman T didistribusikan ke halaman A.

Contoh:



- PageRank ( $PR$ ) of page  $C = PR(A)/2 + PR(B)/1$
- More generally,

$$PR(u) = \frac{\lambda}{N} + (1 - \lambda) \cdot \sum_{v \in B_u} \frac{PR(v)}{L_v}$$

- where  $B_u$  is the set of pages that point to  $u$ , and  $L_v$  is the number of outgoing links from page  $v$  (not counting duplicate links)
- Don't know PageRank values at start
- Assume equal values ( $1/3$  in this case), then iterate:
  - first iteration:  $PR(C) = 0.33/2 + 0.33 = 0.5$ ,  $PR(A) = 0.33$ , and  $PR(B) = 0.17$
  - second:  $PR(C) = 0.33/2 + 0.17 = 0.33$ ,  $PR(A) = 0.5$ ,  $PR(B) = 0.17$
  - third:  $PR(C) = 0.42$ ,  $PR(A) = 0.33$ ,  $PR(B) = 0.25$
- Converges to  $PR(C) = 0.4$ ,  $PR(A) = 0.4$ , and  $PR(B) = 0.2$
- Taking random page jump into account,  $1/3$  chance of going to any page when  $r < \lambda$
- $PR(C) = \lambda/3 + (1 - \lambda) \cdot (PR(A)/2 + PR(B)/1)$
- More generally,

$$PR(u) = \frac{\lambda}{N} + (1 - \lambda) \cdot \sum_{v \in B_u} \frac{PR(v)}{L_v}$$

where  $N$  is the number of pages,  $\lambda$  typically 0.15