

Rankning av WEB sidor

Följande ide kan användas för att rangordna betydelsen av webbsidor i sökfrågor. Tänk på Internet, eller någon annan uppsättning objekt med interna referenser, representerade som ett nätverk med riktade länkar. Intuitivt är det uppenbart att strukturen för detta nätverk kan användas för att gissa den relativa betydelsen av olika sidor (noder). Två begrepp är centrala: sidor med många inkommande länkar och sidor med många utgående länkar. Att hitta sådana noder är naturligtvis enkelt. Men utan ytterligare förfining är utgående eller inkommande länkar inte en jätte bra grund för rangordning. Vad vi vill istället är att identifiera noder med många inkommande länkar där länkarna har sitt ursprung i noder med många utgående länkar, eller vice versa. Vi säger att dessa noder har högt *myndighet* (“authority”) eller *nav* (“hub”) värde.

Problemet kan formuleras med hjälp av “adjacency” matrisen (matris definieras som $a_{ij} = 1$ om det finns en länk från nod i till nod j och 0 annars), då gäller

$$Ay = \lambda x \quad A^T x = \mu y, \quad (1)$$

där vektorerna x och y definierar myndighet och nav värde respektive, och λ och μ är skalärer. För att hitta sidor med det högsta myndigheten och nav värdet vi vill först maximera μ och λ och sedan söka efter stora element i x eller y (elementen i respektive vektor kommer att mäta myndighets eller nav värdet för motsvarande noder).

Ditt problem är att skärpa ide skissen ovan och göra den konkret (matematiskt och algoritmiskt). Använd ditt “system” för att identifiera noder med hög auktoritet och nav värde i nätet definierat av en adjacency matris som du ftt.