Written Assignment Four

CSCI 335.01 – Web Information Retrieval

Assigned: Friday, October 13th, 2006
Due: Friday, October 20th, 2006

This assignment is designed for you to get familiar with the concepts of hubs and authorities using the HITS algorithm, and the PageRank algorithm.

Given the following four web pages and their reference links, compute the first two rounds the following algorithms. Show the results and the steps of computation. If you use programs or spreadsheets, show the formula and results on a separate piece of paper besides submitting the programs or the spreadsheets to me.

![Figure 1: Configuration of the four web pages](image)

1. The HITS algorithm for hubs and authorities.
   
   Initialize for all $p \in S$: $a_p = h_p = 1$
   For $i = 1$ to numOfSteps do
   For all $p \in S$: $a_p = \sum_{q:p \rightarrow q} h_q$
   For all $p \in S$: $h_p = \sum_{q:p ightarrow q} a_q$
   For all $p \in S$: $a_p = \frac{a_p}{c}$ where $\sum_{p \in S}(a_p/c)^2 = 1$
   For all $p \in S$: $h_p = \frac{h_p}{c}$ where $\sum_{p \in S}(h_p/c)^2 = 1$

2. The PageRank algorithm for page ranks.
   
   Let $S$ be the total set of pages.
   Let $\forall p \in S: E(p) = \alpha/|S|$ (e.g. $\alpha = 0.15$)
   Initialize $\forall p \in S: R(p) = 1/|S|
   Until ranks do not change (much)
   for each $p \in S$:
   $R'(p) = \sum_{q:p \rightarrow q} \frac{R(q)}{N_q} + E(p)$
   $c = \frac{1}{\sum_{p \in S} R'(p)}$
   for each $p \in S$: $R(p) = c * R'(p)$