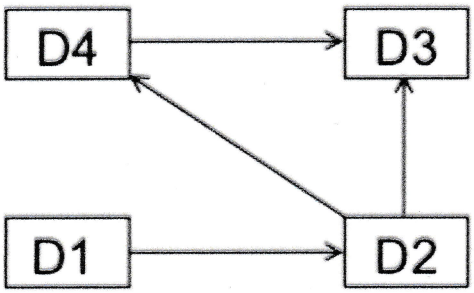
Martyna Stasiak id. 156071

INFORMATION RETRIEVAL – SHORT EXERCISES IV – HITS, RELEVANCE FEEDBACK AND SPELLING CORRECTION

I. Consider the web graph presented below to the left. It involves four pages D1-D4 and four links. Fill in the adjacency matrix L given to the right.



	D1	D2	D3	D4
D1	0	1	0	0
D2	0	0	1	1
D3	0	0	0	0
D4	0	0	1	0

hub authority

The principal eigenvector of LL^T is [0, 1.618, 0, 1] and the principal eigenvector of L^TL is [0, 0, 1.618, 1].

What is $h(D_4)$? Answer: 1 -> $h(D4) = \mu \cdot v \cdot \underset{(D4)}{L \cdot L^T} \cdot h(D4)$ -> $h(D4) = \mu \cdot v \cdot 1 \cdot h(D4)$ -> $\mu \cdot v = 1$ so $h(D4) = 1$

The page with the greatest authority score is: Doc 3

II. Compute the Levenshtein distance for "LEGIA" and "LECHIA".

		L	E	C	H	I	A
	0	1	2	3	4	5	6
L	1	0	1	2	3	4	5
E	2	1	0	1	2	3	4
G	3	2	1	1	2	3	4
I	4	3	2	2	2	2	3
A	5	4	3	3	3	3	2

final distance: 2

- diagonal upper-left
  - for copying characters -> cost 0
  - for substituting character -> cost 1
- left (for insertion) -> cost 1
- top (for deletion) -> cost 1