Mastennikova Aleksandra matricola: 590380

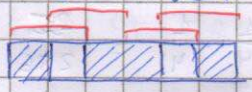
Ex 2.

S = {abaco, adagio, baco, alano}

2-gram index:

$a → abaco, adagio, alano
ab → abaco
ba → abaco, baco
ac → abaco, baco
co → abaco, baco
ad → adagio
da → adagio
ag → adagio
gi → adagio
io → adagio
$b → baco
al → alano
la → alano
an → alano
no → alano

We divide our query strings in 2-grams and we unite the inverted lists of this 2-grams, remembering in how many lists each word occured (we can do it in one scan). After that we select only those words which occured at least |q|-2·e, where |q| is the length of the query string (and number of k-grams it produces) and e is number of errors we permit. The formula is like this because if we ommit errors we can't expect that all the k-grams will occur in our potential candidate; e·2 k-grams won't occur in it



1) p_2 = abco $|p_2|=4$ $e=1$ $\geq 4-2=2$

$a → abaco, adagio, alano
ab → abaco
bc →
co → abaco, baco

=> candidates = <abaco, 3>, <adagio, 1>, <alano, 1>, <baco, 1>
↓ ≥2
"good candidates" = abaco

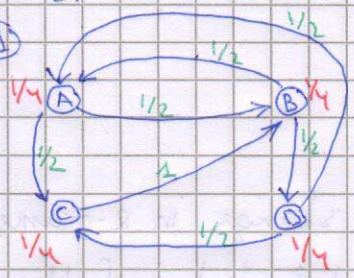
2) p_2 = acabi $|p_2|=5$ $e=1$ $\geq 5-2=3$

$a → abaco, adagio, alano
ac → abaco, baco
ca →
ab → abaco
bi →

=> candidates = <abaco, 3>, <adagio, 1>, <alano, 1>, <baco, 1>
↓ ≥3
"good candidates" = abaco

Ex. 3.

①



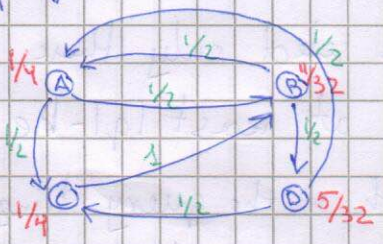
step 1:

$$\text{pagerank}(A) = \frac{3}{4}\left(\frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{2}\right) + \frac{1}{4}\cdot\frac{1}{4} = \frac{4}{16} = \frac{1}{4}$$

$$\text{pagerank}(B) = \frac{3}{4}\left(\frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot 1\right) + \frac{1}{4}\cdot\frac{1}{4} = \frac{3}{4}\cdot\frac{3}{8} + \frac{1}{16} = \frac{11}{32}$$

$$\text{pagerank}(C) = \frac{3}{4}\left(\frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{2}\right) + \frac{1}{4}\cdot\frac{1}{4} = \frac{4}{16} = \frac{1}{4}$$

$$\text{pagerank}(D) = \frac{3}{4}\left(\frac{1}{4}\cdot\frac{1}{2}\right) + \frac{1}{4}\cdot\frac{1}{4} = \frac{3}{32} + \frac{1}{16} = \frac{5}{32}$$



step 2:

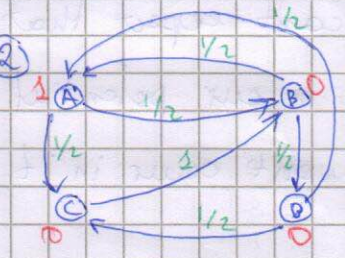
$$\text{pagerank}(A) = \frac{3}{4}\left(\frac{11}{32}\cdot\frac{1}{2} + \frac{5}{32}\cdot\frac{1}{2}\right) + \frac{1}{4}\cdot\frac{1}{4} = \frac{3}{4}\cdot\frac{16}{64} + \frac{1}{16} = \frac{4}{16} = \frac{1}{4}$$

$$\text{pagerank}(B) = \frac{3}{4}\left(\frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot 1\right) + \frac{1}{4}\cdot\frac{1}{4} = \frac{11}{32}$$

$$\text{pagerank}(C) = \frac{3}{4}\left(\frac{1}{4}\cdot\frac{1}{2} + \frac{5}{32}\cdot\frac{1}{2}\right) + \frac{1}{4}\cdot\frac{1}{4} = \frac{3\cdot 13}{4\cdot 64} + \frac{1}{16} = \frac{39+16}{256} = \frac{55}{256}$$

$$\text{pagerank}(D) = \frac{3}{4}\left(\frac{11}{32}\cdot\frac{1}{2}\right) + \frac{1}{4}\cdot\frac{1}{4} = \frac{33}{256} + \frac{1}{16} = \frac{49}{256}$$

②



step 1: (Personalized PageRank for the node A)

$$\text{perPagerank}(A) = \frac{1}{2}\left(\frac{1}{2}\cdot 0 + 0\cdot\frac{1}{2}\right) + \frac{1}{2}\cdot 1 = \frac{1}{2}$$

$$\text{perPagerank}(B) = \frac{1}{2}\left(1\cdot\frac{1}{2} + 0\cdot 1\right) + \frac{1}{2}\cdot 0 = \frac{1}{4}$$

$$\text{perPagerank}(C) = \frac{1}{2}\left(1\cdot\frac{1}{2} + 0\cdot\frac{1}{2}\right) + \frac{1}{2}\cdot 0 = \frac{1}{4}$$

$$\text{perPagerank}(D) = \frac{1}{2}\left(0\cdot\frac{1}{2}\right) + \frac{1}{2}\cdot 0 = 0$$

Ex.