

A.) Vocabulary = { natural, language, processing, becoming, important, since, soon, begin, talking, computer, if, understand, become, much, simpler, use, speech, recognition, first, step, build, like, u}

B.) No. of bigrams = 8

No. of trigrams = 7

C.) [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1]

D.) Cosine similarity : $\cos(x, y) = (x \cdot y) / (|x| |y|)$

$$x \cdot y = 0.2 \cdot 0.3 + 0.1 \cdot 0 + 0.3 \cdot 0.2 + 0.4 \cdot 0.5 = 0.32$$

$$|x| = \sqrt{.04 + .01 + .09 + .16} = \sqrt{0.3}$$

$$|y| = \sqrt{0.09 + 0 + .04 + .25} = \sqrt{0.38}$$

$$\cos(x, y) = (0.32) / (\sqrt{0.3} * \sqrt{0.38}) = 0.9478$$