W1D5

Q1: Write an in-mapper combiner algorithm modifying Co-occurrence Matrix (Pairs approach) algorithm.

Q2: Write an in-mapper combiner algorithm modifying Co-occurrence Matrix (Stripe approach)

Q3: Assume that there are two input splits and two reducers. Note that Mapper 1 and Reducer 1 run on the same machine. Mapper 2 and Reducer 2 run on the same machine.

Further, let the partitioner assign all words less than letter ‘K’ to Reducer 1 and everything else to Reducer 2.

Input Split 1: [{cat mat rat, cat}, { cat bat cat pat}, {cat bat rat bat}] (Note: 3 records)

Input Split 2: [{cat rat bat rat}, {bat mat pat bat}, {pat cat bat mat}] (Note: 3 records)

Let the window of X, W(X) be set of all term after X and before the next X.

Example: Let Data block be [a b c a d e]

W(a) = {b, c}, W(b) = {c, a, d, e}, W(c) = {a, d, e}, W(a) = {d, e}, W(d) = {e}, W€ = {}

1. Illustrate Pair approach

2. Illustrate In-Mapper Combining Version of the Pair approach. (The algorithm you wrote in Q1)

3. Illustrate Stripe approach.

4. Illustrate In-Mapper Combining Version of the Stripe approach. (The algorithm you wrote in Q2)