

W2D2 Solution Q1a: Build inverted index using the "Modified" Algorithm.

INPUT	Input Split-1	Input Split-2	Input Split-3
Mapper Input	<div>DocID 101</div> <div>cat pat mat sat cat eat</div>	<div>DocID 201</div> <div>pat mat sat pat mat eat</div>	<div>DocID 301</div> <div>sat mat cat pat fat mat</div>
MAP	Mapper-1	Mapper-2	Mapper-3
Mapper Output	((cat,101), 2) ((pat,101), 1) ((mat,101), 1) ((sat,101), 1) ((eat,101), 1)	((pat,201), 2) ((mat,201), 2) ((sat,201), 1) ((eat,201), 1)	((sat,301), 1) ((mat,301), 2) ((cat,301), 1) ((pat,301), 1) ((fat,301), 1)
PARTITION	(cat, sat) -> Reducer-1	(others) -> Reducer-2	
	((cat,101), 2) ((sat,101), 1) ((sat,201), 1) ((sat,301), 1) ((cat,301), 1)	((pat,101), 1) ((mat,101), 1) ((eat,101), 1) ((pat,201), 2) ((mat,201), 2) ((eat,201), 1) ((mat,301), 2) ((pat,301), 1) ((fat,301), 1)	
SHUFFLE & SORT			
Reducer Input	((cat,101), [2]) ((cat,301), [1]) ((sat,101), [1]) ((sat,201), [1]) ((sat,301), [1])	((eat,101), [1]) ((eat,201), [1]) ((fat,301), [1]) ((mat,101), [1]) ((mat,201), [2]) ((mat,301), [2]) ((pat,101), [1]) ((pat,201), [2]) ((pat,301), [1])	

REDUCE	Reducer-1	Reducer-2
<p>Note:</p> <pre> class Reducer { initialize() reduce(tuple<t,n>,tf[f]) close() } </pre>	<pre> tprev ≠ ∅ P ← new PostingsArray t=cat (cat ≠ ∅):true (∅ ≠ ∅):false P ← {101:2} tprev=cat t=cat (cat ≠ cat):false P ← {101:2} → {301:1} tprev=cat t=sat (sat ≠ cat):true (cat ≠ ∅):true (cat, {101:2} → {301:1}) P ← new PostingsArray P ← {101:1} tprev=sat t=sat (sat ≠ sat):false P ← {101:1} → {201:1} tprev=sat t=sat (sat ≠ sat):false P ← {101:1} → {201:1} → {301:1} tprev=sat (sat, {101:1} → {201:1} → {301:1}) </pre>	<pre> tprev ≠ ∅ P ← new PostingsArray t=eat (eat ≠ ∅):true (∅ ≠ ∅):false P ← {101:1} tprev=eat t=eat (eat ≠ eat):false P ← {101:1} → {201:1} tprev=eat t=fat (fat ≠ eat):true (eat ≠ ∅):true (eat, {101:1} → {201:1}) P ← new PostingsArray P ← {301:1} tprev=fat ... t=pat (pat ≠ pat):false P ← {101:1} → {201:2} tprev=pat t=pat (pat ≠ pat):false P ← {101:1} → {201:2} → {301:1} tprev=pat (pat, {101:1} → {201:2} → {301:1}) </pre>
Reducer Output	<pre> (cat, {101:2} → {301:1}) (sat, {101:1} → {201:1} → {301:1}) </pre>	<pre> (eat, {101:1} → {201:1}) (fat, {301:1}) (mat, {101:1} → {201:2} → {301:2}) (pat, {101:1} → {201:2} → {301:1}) </pre>