



ASSIGNMENT 4 REPORT

Marrium Jilani – 20K-1748



QUESTION #1

MAPPING

- The WordCountMapper will map the input using key-value pairs with the year and the journal/booktitle as the key and 1 as the value.
- In the map stage, the year and the journal/booktitle will be extracted from each record in the output.csv file and then key-value pairs will be made.

REDUCING

- In the reduce stage, a key-value pair will be received with the year and the journal/booktitle as the key and a list of values consisting of 1s.
- The WordCountReducer will sum up the list of 1s to obtain the total number of articles published in the journal in the given year.
- Finally, it will generate key-value pairs with the number of articles published in that year as the value.

SCREENSHOTS:

```
1997,Digital System Research Center Report      1
1997,Simul.          66
1997,Universit      27
1998,Simul.          87
1998,Universit      30
1999,IWBS Report      1
1999,Inf. Knowl. Syst. Manag.      18
1999,Simul.          94
1999,Universit      29
2000,Inf. Knowl. Syst. Manag.      12
2000,J. Econ. Theory      73
2000,Simul.          71
2000,Universit      13
2001,Inf. Knowl. Syst. Manag.      12
2001,J. Econ. Theory      89
2001,Manag. Sci.      115
2001,Simul.          108
2001,Universit      19
2002,Inf. Knowl. Syst. Manag.      5
2002,J. Econ. Theory      128
2002,Manag. Sci.      104
2002,Simul.          64
2002,Universit      23
2003,Comput. Chem. Eng.      137
2003,Inf. Knowl. Syst. Manag.      7
2003,Internet Math.      20
2003,J. Econ. Theory      109
2003,Manag. Sci.      112
2003,Simul.          49
2003,Universit      6
```

1988,LILOG-Report	29	
1988,Simul.	71	
1989,GTE Laboratories Incorporated		1
1989,IWBS Report	34	
1989,LILOG-Memo	1	
1989,Simul.	90	
1990,ANSI X3H2	2	
1990,GTE Laboratories Incorporated		1
1990,IWBS Report	38	
1990,Simul.	97	
1991,ANSI X2H2	1	
1991,ANSI X3H2	1	
1991,GTE Laboratories Incorporated		3
1991,IWBS Report	53	
1991,Simul.	125	
1992,ANSI X3H2	1	
1992,IWBS Report	32	
1992,Simul.	114	
1992,Universit	28	
1993,GTE Laboratories Incorporated		4
1993,Simul.	100	
1993,Universit	24	
1994,GTE Laboratories Incorporated		1
1994,Simul.	87	
1994,Universit	18	
1995,Compulog Newsletter	1	
1995,GTE Laboratories Incorporated		1
1995,Simul.	86	
1995,Universit	24	
1996,"ETH Zurich, Department of Computer Science / Technical Report"		1

A few screenshots of the output are attached because the output was huge.

1989, Simul.	90
1990, ANSI X3H2	2
1990, GTE Laboratories Incorporated	
1990, IWBS Report	38
1990, Simul.	97

16	1979	UCB/ERL M79/28	University of California at Berkeley
17	1990	X3H2-90-412	ANSI X3H2
18	1991	DBL:KAW-006 X3H2-91-133rev1	ANSI X2H2
19	1990	X3H2-90-292	ANSI X3H2
20	1991	DBL:ARL-029 X3H2-91-083rev1	ANSI X3H2
21	1992	X3H2-92-062	ANSI X3H2
22	1991		191 IWBS Report
23	1988		59 LILOG-Report
24	1987		15 LILOG-Report
25	1988		40 LILOG-Report
26	1992		225 IWBS Report
27	1987		2 LILOG-Memo
28	1988		61 LILOG-Report
29	1988		53 LILOG-Report
30	1988		9 LILOG-Memo
31	1988		45 LILOG-Report
32	1988		56 LILOG-Report
33	1991		201 IWBS Report

QUESTION #2

Pseudo-Code:

Mapping:

1. for each line in the output.csv file:
2. extract the list of authors from the second column
3. sort the list of authors alphabetically to avoid duplicate pairs
4. for each pair of authors in the list:
5. generate (Text(authorpair), IntWritable(1))

Reducing:

1. for each authorpair, counts in the output.csv file:
2. sum up the counts for the authorpair
3. generate (Text(authorpair), IntWritable(sum))

CoAuthorshipMapper:

This function will read each line from the output.csv file, extract the list of authors from the second column, and generates all pairs of authors who have published articles together. The key of the output (key, value) pair is a Text object representing the pair of authors (in alphabetical order to avoid duplicate pairs), and the value is an IntWritable object with a value of 1, representing the publication count.

CoAuthorshipReducer:

This function will receive pairs of authors and their corresponding publication counts generated by the mapper function. The reducer function will sum up the publication counts for each pair of authors and generate the final (key, value) pair, where the key is a Text object representing the pair of authors, and the value is an IntWritable object representing the total number of publications.

OUTPUT:

Here we can see that the pair of authors below have co-authored 2 publications together which will be the weight on the graph edge and that key which is the pair will be the node.

```
Yufei Wu 0014,Zihan Yan 1
Yugeng Xi 0001,Zolt 2
Yuhan Ding,Zongyi Sun 1
Yuhang Xing,Zhiyao Chen 1
Yuhua Liu,Zhao Xiao 1
Yuhua Liu,Zhiqiang Zhou 2
```

7400	[Tariq Boukrouya, 'Vasilios Plets', 'Nikhil Nathacharandit', 'Fernando J. Muzzio', 'Marina M. C. Terapentou']
7407	[Qun Wu', 'Yugeng Xi 0001', 'Zolt', 'Dewei Li']
7408	['Julien Billeter', 'Diogo Rodrigues', 'Sriniketh Srinivasan', 'Michael Amrhein', 'Dominique Bonvin']
7409	['Sebastian C. Brandt', 'Jan Morbach', 'Michalis Miatidis', 'Manfred Thei', 'Matthias Jarke', 'Wolfgang Marquardt']
7410	['Yingjie Ma', 'Yiqing Luo', 'Shuo Zhang', 'Xigang Yuan']
7411	['Woosuk Sung', 'Jaewook Lee 0004']
7412	['Vladim', 'Petr Koc', 'Frantisek Step', 'Milan Kubicek', 'Milos Marek']
7413	['Luis F. Dom', 'Efstratios N. Pistikopoulos']
7414	['Haijuan Cui', 'Xiaochuan Luo']
7415	['Aaron Spettl', 'Maksym Dosta', 'Frederik Klingner', 'Stefan Heinrich', 'Volker Schmidt']
7416	['Junwen Luo', 'Chunjian Xu', 'Yawen Zhang', 'Kaixin Yan', 'Jesse Zhu']
7417	['Mingliang Wang', 'Riccardo Sven Risuleo', 'Elling W. Jacobsen', 'V', 'H']
7418	['Tzu-Ming Yeh', 'Ming-Chieh Huang', 'Chi-Tsung Huang']
7419	['Pedro M. Castro', 'Bruno Cust', 'Henrique A. Matos']
7420	['Enso Ikonen', 'Istv', 'Kaddour Najim']

output

Yugeng Xi 0001

Find All ☐ Formatted Display ☐ Match Case

8725	['Qun Wu', 'Yugeng Xi 0001', 'Zolt', 'Dewei Li']
8726	['Satish R. Inamdar', 'I. A. Karimi', 'S. J. Parulekar', 'B. D. Kulkarni']
8727	['Solomon Brown', 'Eric S. Fraga', 'Haroun Mahgerefteh', 'S. Martynov']
8728	['ChangKyoo Yoo', 'Su Whan Sung', 'In-Beum Lee']
8729	['Swati Mohanty']
8730	['P. Cruz', 'J. C. Santos', 'Fern', 'Ad']
8731	['Miguel J. Bagajewicz', 'Andr']
8732	['Naveed Ramzan', 'Muhammad Faheem 0001', 'Rafiqul Gani', 'Werner Witt']
8733	['Balram Suman', 'Nazish Hoda', 'Shweta Jha']
8734	['Tanuj Karia', 'Claire S. Adjiman', 'Benó']
8735	['Esmail Jahanshahi', 'Krishnamoorthy Dinesh 0002', 'Andr', 'Bjarne Foss', 'Sigurd Skogestad']
8736	['Z. Dehouche', 'B. Bernauer', 'J. Lieto']
8737	['Han Wang', 'Richard Mastragostino', 'Christopher L. E. Swartz']
8738	['Asmaa A. R. Awad', 'Hedia Fgaier', 'Ibrahim Mustafa', 'Ali Elkamel', 'Said S. E. H. Elnashaie']

<<
<
>
>>

✖

⌵
↑
↓
Find All
☐ Formatted Display
☐ Match Case

Pavlos Delias,Zoran Obradovic	1	
Pavlos Kotidis,Stefania Trepekli		1
Pavlos Kotidis,Styliani Avraamidou		1
Pavol Kral,Tatiana Corejova	1	
Pavol N,Viera Rozinajov	1	
Pavol Tanuska,Tomas Skripcak	1	
Pavol Tanuska,Uwe Konrad	1	
Pawan Goyal 0003,Peter Benner	2	
Pawel Dlotko,Victor M. Zavala	1	
Pawel Gburzynski,Weiguang Shi	1	
Pawel Pralat,Przemyslaw Szufel	1	
Pawel Pralat,Rory Wilson	1	
Pawel Pralat,Tomasz Luczak 0001	1	
Pawel Pralat,Vittorio Scarano	1	
Pawel Pralat,William Aiello	1	
Pawel Pralat,Xavier P	1	
Pawel Sztromwasser,Sattanathan Subramanian		1
Payala Venkat Vikash,Yogendra N. Shastri		1
Payman Salamati,Zohrehsadat Naji	2	
Pedro ,Sandra Baldassarri	1	
Pedro ,Teresa Coma-Rosello	1	
Pedro ,Valentina Nisi	1	
Pedro A. Castillo-Castillo,Vladimir Mahalec		2
Pedro A. Ferreira,Rahul Telang	3	
Pedro A. Ferreira,Rodrigo Belo	2	
Pedro Alvarez,Xiuchao Wu	1	
Pedro Alves Nogueira,Rui Rodrigues 0001	1	
Pedro Amorim,T	1	
Pedro C. Pautasso,Vanina G. Cafaro	1	
Pedro C. Pautasso,Victor M. Zavala	1	

Pedro H. R. Peixoto,Wellington Balmant	1		
Pedro Henrique Borges,Rener Martins Moura		1	
Pedro Henrique Borges,Rubens Gedraite	1		
Pedro Henrique Triguís Schimit,Sidnei Alves de Ara			1
Pedro Javier Navarro Lorente,Pedro M. Alcover	1		
Pedro L. D. Peres,Ricardo C. L. F. Oliveira	3		
Pedro M. Castro,Peter Bongers	1		
Pedro M. Castro,Qi Liao 0001	2		
Pedro M. Castro,Qi Zhang 0016	1		
Pedro M. Castro,Sebastian Engell		1	
Pedro M. Castro,Sebastian Forssell		1	
Pedro M. Castro,Susana Relvas	1		
Pedro M. Castro,Yamin Yan	1		
Pedro M. Castro,Yongtu Liang	2		
Pedro M. Saraiva,Pedro V. Lima	1		
Pedro M. Saraiva,Rodolfo Oliveira		1	
Pedro M. Saraiva,Tiago Dias	1		
Pedro Manuel Sasia Santos,Susana Romero Yesa		1	
Pedro Mar,Ruth Herrero	1		
Pedro Morillo 0001,Ramiro Serrano Vergel		1	
Pedro Morillo 0001,Sergio Casas-Yrurzum	1		
Pedro Morillo Tena,Ramiro Serrano Vergel		1	
Pedro Morillo Tena,Sergio Casas-Yrurzum	1		
Pedro Pereira 0002,Tiago Pires	1		
Pedro Rivotti,Romain S. C. Lambert	2		
Pedro Santos Bartolom,Tom Van Gerven	1		
Pedro X. La Hera,Simon Westerberg	1		
Peep K,Shahab Mokarizadeh	1		
Peer-Olaf Siebers,Uwe Aickelin	1		