Part 1.
Question 1:

SELECT ROW_NUMBER() OVER (order by num desc), title, p, pages ,num from
(
select title, publisher as p , pages, COUNT(*) as num from Book1 b1
group by title

UNION ALL

select book_title as title, publisher_name as p, pages ,COUNT(*) as num from Book2 b2
group by book_title

UNION ALL

select Title as title, Publisher as p, Pages as pages ,COUNT(*) as num from Book3 b3
group by Title

UNION ALL

select Title as title, Publisher as p , Pages as pages, COUNT(*) as num from Book4 b4
group by Title
)
 limit 100;

Question 2:

2.1

Mixed Fragmentation should be used to fragment Book2 table.
Predicates Pr = {publisher_name="XXX1",publisher_name="XXX2"… publisher_name="XXXn",,
publication_year=="the given year", publication_year!="the given year"}
and schema divided : [id, book_title, authors, language, pages] and [publication_year,
publication_maonth, publication_day, edition, publisher_name, isbn13, series]

If only 2 fragments are generated, I will use Vertical fragment,
the schema is divided into :
[id, book_title, authors, language, pages] and [publication_year, publication_maonth,
publication_day, edition, publisher_name, isbn13, series]

2.2

not set valid , because is:Disjointness is not fulfilled, as Fragement 2 and Fragent 3 have
jointness.

min-term predicate steps:
step1:
p1 : pages ≤ 100
p2 : pages ≥ 100 and pages ≤ 800
p3:   pages ≥ 600
step2:
Pr = {p1,p2, p3} which is not Disjointness, Minterm predicates:
m1 : (pages ≤ 100)
m2: (pages ≥ 600)

m3 : NOT(pages ≤ 100) and NOT(pages3 ≥ 600)=(pages >100) and(pages < 600)=(100<page<600)
then switch m2 and m3 places:
m1 : (pages ≤ 100)
m2: (100<page<600)
m3: (pages ≥ 600)
step3:
then outcome of M = { m1, m2, m3} ,where m1 : (pages ≤ 100) m2: (100<page<600) m3: (pages ≥ 600)

if we want to insert a new record into Book2:
1. check new insert data's pages, to know new record should insert to m1 range or m2 range or m3 range
2. in the range the new record should be inserted, check the fragment table's status and space, whether the new record can/be inserted, if we can , the insert into it , if now , we should do shrink table and make space for the new record
additon: if we are with and above Oracle11g, because we are range fragment, once a new record in insert, we could use "Interval partition" ,Oracle wiill automatically manag partition.

Part 2

Question 3:

Fact table has the most records.
Because Fact table associated with each fact is a key identifying, for example, which day, which book and which publisher and which language.

Question 4

4.1

The advantages of using bitmap indexes are：
Significant reduction in space and I/O
Reduce processing time：

        Comparison, join and aggregation operations can be reduced to bit operation ；
        Bit operations are very fast

Bitmap indexs is not suitable for high cardinality domains

4.2

publisher	language
AAAI Press | Springer International Publishing | Springer London |IEEE Computer Society Press	English |Spanish

publisher	language
1 0 0 0	1 0
0 1 0 0	1 0
0 0 1 0	1 0
0 0 0 1	1 0
1 0 0 0	0 1
0 1 0 0	0 1
0 0 1 0	0 1
0 0 0 1	0 1

then the bitmap index is :

```
1 0 0  0  1 0
0 1 0  0 1 0
0 0 1 0  1 0
0 0  0 1 1 0
1 0 0  0 0 1
0 1 0  0 0 1
0 0 1 0   0 1
0 0  0 1 0 1
```

4.3

bitmap of "AAAI Pres" and language "english" is: 100010
In order to find the total sales of "English" books published by "AAAI Press":
we can filter out **T** from fact tables by bitmap index = 100010
then Select   SUM(Sales) from **T**

Part 3

Question 5

5.1

global conceptual schema is:
FullBook(id, title, authors, publish_date, publisher, isbn13, pages)

5.2

code is at: Assignment/src/Q5_2Integrate_Book3Book4.py
output csv is at: Assignment/out/FullBookList.csv

1	1757	Representing Femininity: Middle-Class Subjectivity in Victor	['Mary Jean Corbett', '', '']	01/28/1992	Oxford University Press	2900195068589	
2	2133	Autobiography of Yukichi Fukuzawa	['Yukichi Fukuzawa', '', '']	01/02/2007	Columbia University Press	2900231139860	480
3	2862	Living of Charlotte Perkins Gilman: An Autobiography	['Charlotte Perkins Gilman', 'Zona Gale', '']	02/28/1991	University of Wisconsin Press	2900299127441	394
4	2265	Caribbean Autobiography: Cultural Identity and Self-Repres	['Sandra Pouchet Paquet', '', '']	01/28/2002	University of Chicago Press	2900299176944	
5	3080	Living Our Stories, Telling Our Truths: Autobiography and th	['V. P. Franklin', 'Vincent Paul Franklin', '']	01/28/1996	Simon & Schuster Adult Publishing Group	2900689121929	
6	1451	The Education of Fanny Lewald: An Autobiography / Edition	['Fanny Lewald', 'Hanna Ballin (Ed. and Trar	11/03/1992	State University of New York Press	2900791411482	341
7	2939	Mary: An Autobiography / Edition 1	['Mary E. Mebane', '', '']	03/22/1999	The University of North Carolina Press	2900807848219	256
8	3273	Reading Autobiography: A Guide for Interpreting Life Narrat	['Sidonie Smith', '', '']	01/01/2001	University of Chicago Press	2900816628833	
9	1657	Overturned Chariot: The Autobiography of Phan-Boi-Chau	['Phan Boi Chau', 'Boi Chau Phan', '']	01/01/1999	University of Hawaii Press, The	2900824818752	
10	1619	Afrika My Music: An Autobiography 1957-1983	['Ezekiel Mphahlele', '', '']	12/01/1995	Ohio University Press	2900869754848	260
11	3413	Afrika My Music: An Autobiography 1957-1983	['Ezekiel Mphahlele', '', '']	12/01/1995	Ohio University Press	2900869754848	260
12	1360	Havana: Autobiography of a City	['Alfredo Jose Estrada', '', '']	03/18/2008	St. Martin's Press	2901403975392	288
13	1695	Life Under Slavery: Autobiographies of Three American Slav	['Lenny Flank', '', '']	11/10/2009	Lenny Flank	2940000721483	
14	2140	Life Under Slavery: Autobiographies of Three American Slav	['Lenny Flank', '', '']	11/10/2009	Lenny Flank	2940000721483	
15	981	Autobiography of a Pocket Handkerchief	['James Fenimore Cooper', '', '']	09/01/2009	B&R Samizdat Express	2940000735459	
16	649	A Burlesque Autobiography	['Mark Twain', '', '']	09/01/2009	B&R Samizdat Express	2940000737040	
17	334	Chapters from my Autobiography	['Mark Twain', '', '']	02/26/2012	B&R Samizdat Express	2940000737101	
18	333	The Confessions of St. Augustine	['Saint Augustine', '', '']	09/01/2009	B&R Samizdat Express	2940000739426	
19	1040	Autobiography: Truth and Fiction Relating to My Life	['Johann Wolfgang von Goethe', '', '']	12/01/2008	B&R Samizdat Express	2940000749593	
20	2612	James Nasmyth: Engineer, an Autobiography	['James Nasmyth', '', '']	01/01/2009	B&R Samizdat Express	2940000764077	

Question 6

code is at: Assignment/src/Q6.py
the linkage-pair file we generated is at: Assignment/out/b1_b2_linkage.csv

tp, fp, fn= 212 937 20
precision= 0.18450826805918188
recall= 0.9137931034482759
f_measure= 0.3070238957277335

Question 7

code is at: Assignment/src/Q7.py

field_count= 62900 error_count= 28633
error_count / field_count= 0.455214626391097