

Al Albi
Dong Shin
CSE 4701
April 1, 2018
Project 2

The following uses JDBC on Java, and was compiled in an Eclipse IDE.

This second project involves some data mining from a genetic dataset. In particular, we were assigned to find these particular details: for any sets in the Query table and in the Gene tables that have genes in common, we display the number of genes that they share, as well as the number of genes that each individual set has in total. Once those details were obtained, they would be printed as a table.

To implement this project, I took advantage of Java's object oriented structure to organize the information retrieved from SQL queries. In particular, I created a data structure that stored each identifying gene set being compared to, the number of genes that set has in common with one of the query sets, and the amount of time calculated comparing those two sets in a compact manner. I also created multiple methods to organize several other queries and placed them in their own data structures. After gathering all of these elements in a table, the results are shown.

An interesting observation I made was that JDBC did not feature LIMIT, which would have made it significantly advantageous to isolate the five rows where gene sets and query sets share the most seats. Although I used the number of rows to try and limit this, as well as using ORDER BY, it did not turn the expected results. To overcome this, I decided instead to retrieve the entire inner join of the query and gene sets by their gene id's. As a result, the computation for every query set is significantly higher.

The following is my table for Gene Set 1 and Gene Set 2:

Part A

Table for G1

Query Set ID	Q Size	Gene Set ID	G Size	O CNT	Time
1	258	4790	1991	37	657.0
1	258	5855	1800	34	657.0
1	258	1902	1994	34	657.0
1	258	1989	1480	33	657.0
1	258	1172	1690	32	657.0
2	465	3665	1835	55	632.0
2	465	5667	1880	55	632.0
2	465	4139	1852	53	632.0
2	465	4631	1771	53	632.0
2	465	5946	1974	52	632.0
3	382	5673	1969	50	629.0
3	382	3167	1992	50	629.0
3	382	2254	1984	49	629.0
3	382	2486	1959	46	629.0
3	382	3428	1791	46	629.0
4	389	3475	1984	54	626.0
4	389	5385	1883	45	626.0
4	389	1112	1860	45	626.0
4	389	2137	1703	45	626.0
4	389	2443	1837	45	626.0
5	491	3650	1755	59	639.0
5	491	3512	1966	58	639.0
5	491	1250	1984	57	639.0
5	491	4940	1922	56	639.0
5	491	2496	1998	56	639.0
6	181	5049	1702	26	612.0
6	181	4559	1799	26	612.0
6	181	1283	1998	25	612.0
6	181	1430	1677	25	612.0
6	181	4191	1803	25	612.0
7	392	4896	1996	50	626.0
7	392	1952	1970	48	626.0
7	392	3969	1960	48	626.0
7	392	4799	1885	47	626.0
7	392	1280	1968	46	626.0
8	392	3741	1978	50	632.0
8	392	4907	1996	49	632.0
8	392	5212	1957	48	632.0
8	392	1688	1978	47	632.0
8	392	2802	1982	47	632.0
9	467	3214	1923	56	637.0
9	467	4343	1942	56	637.0
9	467	2550	1972	54	637.0
9	467	4121	1992	54	637.0
9	467	1773	1992	53	637.0
10	468	2972	1892	59	638.0
10	468	3499	1876	56	638.0
10	468	1429	1787	56	638.0
10	468	1332	1943	55	638.0
10	468	4452	1811	55	638.0
11	178	1850	1609	27	613.0
11	178	4799	1885	26	613.0
11	178	3833	1835	25	613.0
11	178	2025	1843	24	613.0
11	178	2756	1908	24	613.0
12	485	5526	1890	63	664.0
12	485	3185	1930	60	664.0
12	485	2159	1966	58	664.0
12	485	2209	1910	56	664.0
12	485	5750	1993	55	664.0
13	138	3897	1675	20	630.0
13	138	1318	1899	20	630.0
13	138	4962	1669	20	630.0
13	138	2880	1953	19	630.0
13	138	5416	1996	19	630.0

14	417	2796	1982	52	632.0
14	417	3042	1823	48	632.0
14	417	4134	1704	48	632.0
14	417	4907	1996	48	632.0
14	417	4065	1963	48	632.0
15	286	1920	1997	41	624.0
15	286	1091	1874	38	624.0
15	286	5521	1996	37	624.0
15	286	2913	1896	37	624.0
15	286	5215	1998	36	624.0
16	154	3923	1557	24	620.0
16	154	3427	1820	24	620.0
16	154	5998	1962	23	620.0
16	154	4569	1833	22	620.0
16	154	2699	1485	22	620.0
17	230	3457	1951	35	633.0
17	230	5557	1869	30	633.0
17	230	1291	1997	30	633.0
17	230	2502	1906	30	633.0
17	230	2839	1977	30	633.0
18	144	5044	1949	23	615.0
18	144	2672	1986	22	615.0
18	144	5416	1996	21	615.0
18	144	4898	1665	21	615.0
18	144	1939	1974	20	615.0
19	108	4234	1516	17	609.0
19	108	2854	1955	17	609.0
19	108	5521	1996	17	609.0
19	108	5946	1974	17	609.0
19	108	4140	1515	17	609.0
20	263	2256	1778	34	661.0
20	263	5595	1988	34	661.0
20	263	3896	1716	34	661.0
20	263	5046	1825	34	661.0
20	263	3145	1751	34	661.0
21	81	4709	1823	14	610.0
21	81	4017	1825	14	610.0
21	81	3274	1794	14	610.0
21	81	5493	1989	13	610.0
21	81	2944	1625	13	610.0
22	115	5774	1943	20	620.0
22	115	4506	1818	19	620.0
22	115	5981	1975	18	620.0
22	115	2197	1990	18	620.0
22	115	1419	1764	18	620.0
23	434	1672	1967	59	634.0
23	434	4151	1887	52	634.0
23	434	1119	1969	52	634.0
23	434	1177	1762	50	634.0
23	434	4013	1959	50	634.0
24	293	2315	1970	36	624.0
24	293	4052	1801	36	624.0
24	293	1298	1685	35	624.0
24	293	1453	1827	35	624.0
24	293	1891	1974	35	624.0
25	319	2538	1855	40	623.0
25	319	4157	1926	39	623.0
25	319	3479	1861	39	623.0
25	319	2025	1843	39	623.0
25	319	1140	1969	38	623.0
26	188	5464	1865	27	610.0
26	188	2160	1646	27	610.0
26	188	5603	1819	27	610.0
26	188	3582	1786	26	610.0
26	188	5507	1802	26	610.0
27	410	2496	1998	52	632.0
27	410	1222	1905	49	632.0
27	410	5758	1974	49	632.0
27	410	1645	1947	48	632.0

27	410	2959	1528	47	632.0
28	173	2918	1939	27	624.0
28	173	4041	1960	26	624.0
28	173	5998	1962	26	624.0
28	173	1908	1970	25	624.0
28	173	5545	1852	25	624.0
29	455	1586	1931	58	636.0
29	455	4725	1784	57	636.0
29	455	2976	1911	54	636.0
29	455	2880	1953	53	636.0
29	455	4141	1929	52	636.0
30	295	1465	1955	37	649.0
30	295	1679	1935	36	649.0
30	295	2651	1663	36	649.0
30	295	2946	1963	35	649.0
30	295	5981	1975	35	649.0
31	474	1653	1951	57	633.0
31	474	4410	1658	57	633.0
31	474	5383	1818	56	633.0
31	474	5585	1934	55	633.0
31	474	1585	1988	53	633.0
32	78	3877	1634	15	604.0
32	78	4608	1879	15	604.0
32	78	3934	1986	15	604.0
32	78	3844	1870	14	604.0
32	78	1974	1969	14	604.0
33	438	4560	1731	53	652.0
33	438	1690	1847	53	652.0
33	438	5199	1967	53	652.0
33	438	1902	1994	52	652.0
33	438	3199	1989	51	652.0
34	55	1719	1990	13	605.0
34	55	1010	1951	13	605.0
34	55	1517	1705	12	605.0
34	55	5293	1865	12	605.0
34	55	3634	1519	11	605.0
35	64	2923	1837	13	610.0
35	64	3143	1958	12	610.0
35	64	4649	1752	12	610.0
35	64	4310	1671	11	610.0
35	64	5507	1802	11	610.0
36	146	1496	1940	25	614.0
36	146	4115	1945	25	614.0
36	146	3929	1975	23	614.0
36	146	1307	1493	23	614.0
36	146	3752	1703	23	614.0
37	263	5396	1674	35	615.0
37	263	5595	1988	34	615.0
37	263	6000	1919	34	615.0
37	263	3569	1906	34	615.0
37	263	2452	1941	33	615.0
38	223	4610	1993	30	613.0
38	223	3622	1959	30	613.0
38	223	2650	1763	30	613.0
38	223	3910	1888	29	613.0
38	223	5319	1946	29	613.0
39	346	1330	1992	45	630.0
39	346	4940	1922	45	630.0
39	346	5534	1898	42	630.0
39	346	1492	1932	42	630.0
39	346	2753	1848	41	630.0
40	118	3764	1892	24	607.0
40	118	1205	1958	20	607.0
40	118	5925	1937	19	607.0
40	118	2769	1785	19	607.0
40	118	5507	1802	19	607.0
41	368	4995	1949	47	632.0
41	368	1222	1905	47	632.0
41	368	1274	1911	46	632.0

41	368	5184	1954	44	632.0
41	368	3963	1951	44	632.0
42	445	5150	1913	54	632.0
42	445	1797	1989	54	632.0
42	445	4204	1734	53	632.0
42	445	2115	1964	52	632.0
42	445	1866	1955	52	632.0
43	184	1330	1992	30	612.0
43	184	2300	1922	26	612.0
43	184	4774	1982	26	612.0
43	184	4782	1669	26	612.0
43	184	1797	1989	25	612.0
44	179	1585	1988	31	665.0
44	179	5860	1982	30	665.0
44	179	2800	1734	25	665.0
44	179	4774	1982	25	665.0
44	179	2672	1986	24	665.0
45	218	4543	1846	33	618.0
45	218	5207	1839	29	618.0
45	218	3204	1855	29	618.0
45	218	3641	1824	29	618.0
45	218	2917	1611	29	618.0
46	375	2640	1952	47	935.0
46	375	4614	1957	46	935.0
46	375	2109	1832	46	935.0
46	375	4217	1904	45	935.0
46	375	3934	1986	45	935.0
47	351	4693	1869	50	635.0
47	351	3963	1951	49	635.0
47	351	1893	1977	45	635.0
47	351	3376	1882	44	635.0
47	351	4326	1986	43	635.0
48	260	3262	1800	35	629.0
48	260	3360	1868	34	629.0
48	260	5233	1986	34	629.0
48	260	5140	1880	33	629.0
48	260	5327	1809	32	629.0
49	448	4858	1947	56	638.0
49	448	5070	1765	55	638.0
49	448	2308	1915	53	638.0
49	448	2120	1979	53	638.0
49	448	4664	1862	52	638.0
50	303	3870	1826	40	625.0
50	303	1275	1899	40	625.0
50	303	5703	1966	39	625.0
50	303	4305	1772	37	625.0
50	303	4457	1926	36	625.0

Part B

Table for G2					
Query Set ID	Q Size	Gene Set ID	G Size	O CNT	Time
1	258	8629	383	13	123.0
1	258	7776	495	12	123.0
1	258	8394	381	12	123.0
1	258	9919	450	12	123.0
1	258	10084	463	12	123.0
2	465	8200	488	20	122.0
2	465	9984	471	19	122.0
2	465	7892	470	18	122.0
2	465	6156	482	18	122.0
2	465	7489	492	18	122.0
3	382	10395	397	19	154.0
3	382	10567	352	18	154.0
3	382	7044	478	16	154.0
3	382	6946	379	16	154.0
3	382	6372	465	15	154.0
4	389	9990	376	17	124.0
4	389	6300	498	17	124.0

4	389	9522	452	16	124.0
4	389	6297	473	15	124.0
4	389	9959	480	15	124.0
5	491	8953	492	20	122.0
5	491	7100	429	19	122.0
5	491	8688	456	18	122.0
5	491	9034	444	17	122.0
5	491	6593	447	17	122.0
6	181	6542	359	11	117.0
6	181	6811	295	11	117.0
6	181	10604	489	10	117.0
6	181	10637	488	9	117.0
6	181	7687	480	9	117.0
7	392	6387	442	17	121.0
7	392	8058	459	17	121.0
7	392	9498	486	17	121.0
7	392	9623	492	17	121.0
7	392	9658	477	16	121.0
8	392	10637	488	16	123.0
8	392	7505	354	16	123.0
8	392	10990	497	16	123.0
8	392	8469	302	15	123.0
8	392	7084	481	15	123.0
9	467	7010	490	20	121.0
9	467	8857	418	18	121.0
9	467	7897	458	17	121.0
9	467	6887	499	17	121.0
9	467	9098	452	17	121.0
10	468	10516	499	18	123.0
10	468	10831	451	17	123.0
10	468	8137	424	17	123.0
10	468	9608	448	17	123.0
10	468	8674	436	17	123.0
11	178	6400	452	11	118.0
11	178	8313	328	9	118.0
11	178	9664	471	9	118.0
11	178	8510	447	9	118.0
11	178	10398	494	9	118.0
12	485	6665	495	20	133.0
12	485	8137	424	20	133.0
12	485	8972	461	19	133.0
12	485	9727	458	19	133.0
12	485	7463	460	19	133.0
13	138	7787	402	10	119.0
13	138	6871	428	9	119.0
13	138	9886	490	8	119.0
13	138	8581	497	8	119.0
13	138	7495	328	7	119.0
14	417	9189	432	20	121.0
14	417	7153	419	17	121.0
14	417	7892	470	17	121.0
14	417	10190	472	16	121.0
14	417	8282	380	16	121.0
15	286	7606	496	16	121.0
15	286	10072	383	15	121.0
15	286	10881	408	15	121.0
15	286	9745	481	14	121.0
15	286	10702	435	14	121.0
16	154	10100	486	10	116.0
16	154	7903	433	9	116.0
16	154	8585	347	9	116.0
16	154	8795	446	8	116.0
16	154	9776	494	8	116.0
17	230	6429	489	12	126.0
17	230	10459	499	12	126.0
17	230	7879	499	12	126.0
17	230	9322	457	11	126.0
17	230	7110	423	11	126.0
18	144	7934	469	9	118.0

18	144	7027	470	9	118.0
18	144	7901	490	8	118.0
18	144	10375	422	8	118.0
18	144	10256	422	8	118.0
19	108	7343	298	8	116.0
19	108	9783	430	8	116.0
19	108	9797	460	8	116.0
19	108	8955	463	7	116.0
19	108	6067	490	7	116.0
20	263	7968	358	14	120.0
20	263	7743	363	13	120.0
20	263	10700	486	12	120.0
20	263	10109	480	12	120.0
20	263	9559	479	12	120.0
21	81	6962	491	8	116.0
21	81	8392	441	7	116.0
21	81	10028	231	6	116.0
21	81	10214	403	6	116.0
21	81	9029	395	6	116.0
22	115	8862	390	7	118.0
22	115	9669	372	7	118.0
22	115	10414	480	7	118.0
22	115	6605	302	7	118.0
22	115	9831	254	7	118.0
23	434	9900	403	18	121.0
23	434	7140	386	17	121.0
23	434	6831	499	17	121.0
23	434	6697	468	16	121.0
23	434	7144	496	16	121.0
24	293	8291	438	17	127.0
24	293	9288	486	14	127.0
24	293	9449	436	14	127.0
24	293	9812	475	14	127.0
24	293	7608	419	13	127.0
25	319	8924	466	15	123.0
25	319	7000	436	14	123.0
25	319	6769	461	14	123.0
25	319	10420	415	13	123.0
25	319	6631	447	13	123.0
26	188	6682	493	11	120.0
26	188	7083	481	10	120.0
26	188	8446	467	10	120.0
26	188	8532	478	9	120.0
26	188	7584	491	9	120.0
27	410	7632	444	19	121.0
27	410	10990	497	18	121.0
27	410	8615	435	17	121.0
27	410	10448	496	17	121.0
27	410	8381	394	16	121.0
28	173	7010	490	10	119.0
28	173	7606	496	10	119.0
28	173	9004	390	9	119.0
28	173	7183	456	9	119.0
28	173	10245	422	9	119.0
29	455	8581	497	18	121.0
29	455	7850	493	18	121.0
29	455	6297	473	18	121.0
29	455	9072	430	17	121.0
29	455	9240	486	17	121.0
30	295	9153	479	15	122.0
30	295	6802	423	14	122.0
30	295	7861	468	14	122.0
30	295	8140	447	14	122.0
30	295	6440	415	13	122.0
31	474	9252	483	22	123.0
31	474	10091	466	20	123.0
31	474	9908	487	20	123.0
31	474	9547	412	19	123.0
31	474	7014	489	18	123.0

32	78	6539	477	6	115.0
32	78	6460	409	6	115.0
32	78	10516	499	6	115.0
32	78	6125	240	6	115.0
32	78	6703	414	6	115.0
33	438	10870	495	19	123.0
33	438	6560	476	19	123.0
33	438	9617	433	17	123.0
33	438	9477	494	17	123.0
33	438	9638	394	17	123.0
34	55	9959	480	6	123.0
34	55	10599	448	5	123.0
34	55	9784	344	5	123.0
34	55	6783	218	4	123.0
34	55	9172	357	4	123.0
35	64	9731	470	6	169.0
35	64	7902	304	5	169.0
35	64	7879	499	5	169.0
35	64	6910	444	5	169.0
35	64	8070	491	5	169.0
36	146	8863	470	10	127.0
36	146	8081	496	9	127.0
36	146	8144	361	9	127.0
36	146	7031	453	9	127.0
36	146	10102	481	9	127.0
37	263	8277	451	13	119.0
37	263	8658	428	13	119.0
37	263	9546	472	12	119.0
37	263	7078	378	12	119.0
37	263	10797	414	12	119.0
38	223	8737	419	12	121.0
38	223	7762	494	11	121.0
38	223	10604	489	11	121.0
38	223	10833	494	11	121.0
38	223	9890	469	11	121.0
39	346	8140	447	16	122.0
39	346	6134	369	15	122.0
39	346	9190	475	15	122.0
39	346	7850	493	15	122.0
39	346	10130	407	14	122.0
40	118	7606	496	8	122.0
40	118	9776	494	8	122.0
40	118	7262	456	8	122.0
40	118	10765	465	8	122.0
40	118	10877	396	8	122.0
41	368	9002	452	20	122.0
41	368	6578	443	16	122.0
41	368	9080	393	15	122.0
41	368	6879	473	15	122.0
41	368	7362	397	14	122.0
42	445	8651	438	18	133.0
42	445	9672	489	18	133.0
42	445	9301	476	18	133.0
42	445	9393	447	17	133.0
42	445	6372	465	17	133.0
43	184	9794	392	10	117.0
43	184	7408	380	9	117.0
43	184	9499	496	9	117.0
43	184	10599	448	9	117.0
43	184	7379	493	9	117.0
44	179	6884	471	11	119.0
44	179	10988	451	10	119.0
44	179	6002	440	10	119.0
44	179	6208	484	10	119.0
44	179	9499	496	9	119.0
45	218	6682	493	12	118.0
45	218	8685	403	11	118.0
45	218	7395	450	10	118.0
45	218	7984	487	10	118.0

45	218	6912	428	10	118.0
46	375	9189	432	17	129.0
46	375	6762	463	16	129.0
46	375	10999	492	16	129.0
46	375	8013	493	16	129.0
46	375	8315	459	16	129.0
47	351	10516	499	16	122.0
47	351	10722	452	15	122.0
47	351	6632	454	15	122.0
47	351	10974	445	14	122.0
47	351	6701	400	14	122.0
48	260	9337	478	13	119.0
48	260	6325	482	13	119.0
48	260	10097	397	12	119.0
48	260	8912	463	12	119.0
48	260	9640	463	12	119.0
49	448	6243	461	20	123.0
49	448	7661	481	18	123.0
49	448	10220	471	18	123.0
49	448	8132	483	18	123.0
49	448	6262	456	18	123.0
50	303	10560	495	14	122.0
50	303	10996	447	14	122.0
50	303	9727	458	13	122.0
50	303	10906	402	13	122.0
50	303	8055	499	13	122.0