

Экспериментальный анализ реализации

Автор: Суханова Анжела

Задание: 3

21 сентября 2020

Комментарии к работе

Замеры проводились на ноутбуке с Ubuntu 20.04, Intel Core i5-7300HQ CPU, 2.50GHz, DDR4 8Gb RAM.

В таблицах 1, 2 и 3 представлены средние значения и дисперсии данных (в секундах), полученных по 5-ти запускам для каждого теста. Все данные для замеров взяты из архива refinedDataForRPQ. Измерения были произведены по всем файлам троек вершина-метка-вершина с названием вида **refinedDataForRPQ/LUBM*/LUBM*.txt** и всем файлам регулярных выражений с названием вида **refinedDataForRPQ/LUBM300/regexes/q*0**. Результаты замеров округлены до 6-ого знака после запятой.

1. Исследование производительности реализации пересечения автоматов

Таблица 1: Результаты замеров времени пересечения графов и его вывода, сек

Граф	Регулярное выражение	Пересечение		Вывод	
		Среднее значение	Дисперсия	Среднее значение	Дисперсия
LUBM300/ LUBM300.txt	LUBM300/regexes/q1_0	0.092394	0.000001	0.000002	$< 10^{-6}$
	LUBM300/regexes/q10_2_0	0.099535	0.000011	0.000033	$< 10^{-6}$
	LUBM300/regexes/q10_3_0	0.148202	0.000047	0.000056	$< 10^{-6}$
	LUBM300/regexes/q10_4_0	0.110979	0.000131	0.000055	$< 10^{-6}$
	LUBM300/regexes/q10_5_0	0.12226	0.000227	0.00007	$< 10^{-6}$
	LUBM300/regexes/q11_2_0	0.118073	0.00011	0.000033	$< 10^{-6}$
	LUBM300/regexes/q11_3_0	0.126982	0.000094	0.000038	$< 10^{-6}$
	LUBM300/regexes/q11_4_0	0.133304	0.000103	0.000051	$< 10^{-6}$
	LUBM300/regexes/q11_5_0	0.187005	0.000549	0.000057	$< 10^{-6}$
	LUBM300/regexes/q_12_0	0.186607	0.000016	0.000075	$< 10^{-6}$
	LUBM300/regexes/q_13_0	0.202306	0.000017	0.000068	$< 10^{-6}$
	LUBM300/regexes/q_14_0	0.235481	0.00017	0.000066	$< 10^{-6}$
	LUBM300/regexes/q_15_0	0.147037	0.000007	0.000052	$< 10^{-6}$
	LUBM300/regexes/q_16_0	0.123395	0.000049	0.000069	$< 10^{-6}$
	LUBM300/regexes/q2_0	0.103487	0.000016	0.00003	$< 10^{-6}$
	LUBM300/regexes/q3_0	0.000187	0.000002	0.000044	$< 10^{-6}$
	LUBM300/regexes/q4_2_0	0.101098	0.000018	0.000033	$< 10^{-6}$
	LUBM300/regexes/q4_3_0	0.107645	0.000094	0.000031	$< 10^{-6}$
	LUBM300/regexes/q4_4_0	0.107385	0.00018	0.000051	$< 10^{-6}$
	LUBM300/regexes/q4_5_0	0.110695	0.000279	0.000061	$< 10^{-6}$
	LUBM300/regexes/q5_0	0.117917	0.000106	0.000054	$< 10^{-6}$
	LUBM300/regexes/q6_0	0.147923	0.000015	0.000028	$< 10^{-6}$
	LUBM300/regexes/q7_0	0.091756	0.000005	0.000031	$< 10^{-6}$
	LUBM300/regexes/q8_0	0.090405	0.000004	0.000015	$< 10^{-6}$
	LUBM300/regexes/q9_2_0	0.101761	0.000005	0.00003	$< 10^{-6}$

	LUBM300/regexes/q9_3_0	0.104804	0.00005	0.000038	$< 10^{-6}$
	LUBM300/regexes/q9_4_0	0.097859	0.000026	0.000046	$< 10^{-6}$
	LUBM300/regexes/q9_5_0	0.100923	0.000016	0.000068	$< 10^{-6}$
LUBM500/ LUBM500.txt	LUBM300/regexes/q1_0	0.153106	0.000013	0.000015	$< 10^{-6}$
	LUBM300/regexes/q10_2_0	0.165039	0.000005	0.00005	$< 10^{-6}$
	LUBM300/regexes/q10_3_0	0.166796	0.000051	0.000055	$< 10^{-6}$
	LUBM300/regexes/q10_4_0	0.180987	0.000213	0.000067	$< 10^{-6}$
	LUBM300/regexes/q10_5_0	0.196807	0.000391	0.000069	$< 10^{-6}$
	LUBM300/regexes/q11_2_0	0.194684	0.000174	0.000028	$< 10^{-6}$
	LUBM300/regexes/q11_3_0	0.208278	0.00019	0.000036	$< 10^{-6}$
	LUBM300/regexes/q11_4_0	0.217885	0.000218	0.00004	$< 10^{-6}$
	LUBM300/regexes/q11_5_0	0.301466	0.000862	0.000058	$< 10^{-6}$
	LUBM300/regexes/q_12_0	0.298946	0.000037	0.000062	$< 10^{-6}$
	LUBM300/regexes/q_13_0	0.335591	0.000031	0.000055	$< 10^{-6}$
	LUBM300/regexes/q_14_0	0.404221	0.000207	0.000073	$< 10^{-6}$
	LUBM300/regexes/q_15_0	0.243532	0.000002	0.000053	$< 10^{-6}$
	LUBM300/regexes/q_16_0	0.202833	0.000078	0.000062	$< 10^{-6}$
	LUBM300/regexes/q2_0	0.168221	0.000021	0.000027	$< 10^{-6}$
	LUBM300/regexes/q3_0	0.26005	0.000001	0.000033	$< 10^{-6}$
	LUBM300/regexes/q4_2_0	0.16163	0.000015	0.000034	$< 10^{-6}$
	LUBM300/regexes/q4_3_0	0.173499	0.000147	0.000043	$< 10^{-6}$
	LUBM300/regexes/q4_4_0	0.176938	0.000332	0.000065	$< 10^{-6}$
	LUBM300/regexes/q4_5_0	0.183186	0.000294	0.000062	$< 10^{-6}$
	LUBM300/regexes/q5_0	0.191181	0.000121	0.000039	$< 10^{-6}$
	LUBM300/regexes/q6_0	0.240063	0.00001	0.00003	$< 10^{-6}$
	LUBM300/regexes/q7_0	0.153677	0.000011	0.00005	$< 10^{-6}$
	LUBM300/regexes/q8_0	0.151636	0.000016	0.000017	$< 10^{-6}$
	LUBM300/regexes/q9_2_0	0.165656	0.000014	0.000034	$< 10^{-6}$
	LUBM300/regexes/q9_3_0	0.172665	0.000069	0.000039	$< 10^{-6}$
	LUBM300/regexes/q9_4_0	0.160415	0.000029	0.000048	$< 10^{-6}$
	LUBM300/regexes/q9_5_0	0.165831	0.000049	0.000062	$< 10^{-6}$
LUBM1M/ LUBM1M.txt	LUBM300/regexes/q1_0	0.308594	0.000031	0.000016	$< 10^{-6}$
	LUBM300/regexes/q10_2_0	0.325498	0.00001	0.000034	$< 10^{-6}$
	LUBM300/regexes/q10_3_0	0.336803	0.000022	0.000057	$< 10^{-6}$
	LUBM300/regexes/q10_4_0	0.367562	0.000934	0.000058	$< 10^{-6}$
	LUBM300/regexes/q10_5_0	0.38685	0.00124	0.000065	$< 10^{-6}$
	LUBM300/regexes/q11_2_0	0.384334	0.000725	0.000035	$< 10^{-6}$
	LUBM300/regexes/q11_3_0	0.399925	0.000671	0.000047	$< 10^{-6}$
	LUBM300/regexes/q11_4_0	0.424831	0.000798	0.000055	$< 10^{-6}$
	LUBM300/regexes/q11_5_0	0.599496	0.002933	0.00006	$< 10^{-6}$
	LUBM300/regexes/q_12_0	0.59411	0.000131	0.000047	$< 10^{-6}$
	LUBM300/regexes/q_13_0	0.664955	0.000087	0.00006	$< 10^{-6}$
	LUBM300/regexes/q_14_0	0.767655	0.000719	0.00007	$< 10^{-6}$
	LUBM300/regexes/q_15_0	0.477787	0.000015	0.000061	$< 10^{-6}$
	LUBM300/regexes/q_16_0	0.398895	0.000268	0.000063	$< 10^{-6}$
	LUBM300/regexes/q2_0	0.332708	0.00004	0.000035	$< 10^{-6}$
	LUBM300/regexes/q3_0	0.495416	0.000001	0.000035	$< 10^{-6}$
	LUBM300/regexes/q4_2_0	0.321146	0.00008	0.000028	$< 10^{-6}$
	LUBM300/regexes/q4_3_0	0.341057	0.000572	0.000044	$< 10^{-6}$
	LUBM300/regexes/q4_4_0	0.359573	0.000921	0.00005	$< 10^{-6}$
	LUBM300/regexes/q4_5_0	0.350766	0.00099	0.000067	$< 10^{-6}$

	LUBM300/regexes/q5_0	0.375079	0.000472	0.000038	$< 10^{-6}$
	LUBM300/regexes/q6_0	0.486057	0.00006	0.000039	$< 10^{-6}$
	LUBM300/regexes/q7_0	0.303036	0.000039	0.000037	$< 10^{-6}$
	LUBM300/regexes/q8_0	0.302426	0.000028	0.000015	$< 10^{-6}$
	LUBM300/regexes/q9_2_0	0.31981	0.000043	0.00003	$< 10^{-6}$
	LUBM300/regexes/q9_3_0	0.329048	0.000177	0.000049	$< 10^{-6}$
	LUBM300/regexes/q9_4_0	0.318341	0.000174	0.000052	$< 10^{-6}$
	LUBM300/regexes/q9_5_0	0.327609	0.000098	0.000061	$< 10^{-6}$
LUBM1.5M/ LUBM1.5M.txt	LUBM300/regexes/q1_0	0.450547	0.000096	0.000016	$< 10^{-6}$
	LUBM300/regexes/q10_2_0	0.498264	0.000043	0.00004	$< 10^{-6}$
	LUBM300/regexes/q10_3_0	0.505113	0.000472	0.000055	$< 10^{-6}$
	LUBM300/regexes/q10_4_0	0.549276	0.002626	0.000073	$< 10^{-6}$
	LUBM300/regexes/q10_5_0	0.597628	0.003472	0.000072	$< 10^{-6}$
	LUBM300/regexes/q11_2_0	0.578498	0.002241	0.000034	$< 10^{-6}$
	LUBM300/regexes/q11_3_0	0.605518	0.002043	0.000045	$< 10^{-6}$
	LUBM300/regexes/q11_4_0	0.650001	0.001775	0.000055	$< 10^{-6}$
	LUBM300/regexes/q11_5_0	0.900504	0.007808	0.000051	$< 10^{-6}$
	LUBM300/regexes/q_12_0	0.880336	0.000307	0.000055	$< 10^{-6}$
	LUBM300/regexes/q_13_0	0.983972	0.000288	0.000072	$< 10^{-6}$
	LUBM300/regexes/q_14_0	1.154281	0.003355	0.000069	$< 10^{-6}$
	LUBM300/regexes/q_15_0	0.743516	0.00001	0.000035	$< 10^{-6}$
	LUBM300/regexes/q_16_0	0.590219	0.000549	0.000053	$< 10^{-6}$
	LUBM300/regexes/q2_0	0.504618	0.000047	0.000034	$< 10^{-6}$
	LUBM300/regexes/q3_0	0.771867	0.000004	0.00005	$< 10^{-6}$
	LUBM300/regexes/q4_2_0	0.484303	0.000386	0.000041	$< 10^{-6}$
	LUBM300/regexes/q4_3_0	0.521045	0.001735	0.000037	$< 10^{-6}$
	LUBM300/regexes/q4_4_0	0.528644	0.002373	0.000053	$< 10^{-6}$
	LUBM300/regexes/q4_5_0	0.532871	0.002625	0.000073	$< 10^{-6}$
	LUBM300/regexes/q5_0	0.555894	0.001107	0.00004	$< 10^{-6}$
	LUBM300/regexes/q6_0	0.722074	0.000229	0.000033	$< 10^{-6}$
	LUBM300/regexes/q7_0	0.453372	0.000087	0.000041	$< 10^{-6}$
	LUBM300/regexes/q8_0	0.45248	0.000072	0.000026	$< 10^{-6}$
	LUBM300/regexes/q9_2_0	0.478108	0.000049	0.000031	$< 10^{-6}$
	LUBM300/regexes/q9_3_0	0.503633	0.000363	0.000031	$< 10^{-6}$
	LUBM300/regexes/q9_4_0	0.485235	0.000467	0.000051	$< 10^{-6}$
	LUBM300/regexes/q9_5_0	0.479841	0.000161	0.00003	$< 10^{-6}$
LUBM1.9M/ LUBM1.9M.txt	LUBM300/regexes/q1_0	0.596327	0.00015	0.000016	$< 10^{-6}$
	LUBM300/regexes/q10_2_0	0.626723	0.000068	0.000036	$< 10^{-6}$
	LUBM300/regexes/q10_3_0	0.647579	0.000905	0.000047	$< 10^{-6}$
	LUBM300/regexes/q10_4_0	0.705253	0.004076	0.000064	$< 10^{-6}$
	LUBM300/regexes/q10_5_0	0.734947	0.005071	0.00008	$< 10^{-6}$
	LUBM300/regexes/q11_2_0	0.756009	0.003982	0.000043	$< 10^{-6}$
	LUBM300/regexes/q11_3_0	0.775336	0.003661	0.000037	$< 10^{-6}$
	LUBM300/regexes/q11_4_0	0.838405	0.002955	0.000052	$< 10^{-6}$
	LUBM300/regexes/q11_5_0	1.162913	0.012914	0.000071	$< 10^{-6}$
	LUBM300/regexes/q_12_0	1.215629	0.000419	0.000057	$< 10^{-6}$
	LUBM300/regexes/q_13_0	1.269452	0.000497	0.000069	$< 10^{-6}$
	LUBM300/regexes/q_14_0	1.502229	0.004025	0.000067	$< 10^{-6}$
	LUBM300/regexes/q_15_0	0.938278	0.000045	0.000047	$< 10^{-6}$
	LUBM300/regexes/q_16_0	0.762891	0.0011	0.00007	$< 10^{-6}$
	LUBM300/regexes/q2_0	0.661107	0.000108	0.000031	$< 10^{-6}$
	LUBM300/regexes/q3_0	0.957598	0.000008	0.000041	$< 10^{-6}$

LUBM300/regexes/q4_2_0	0.624364	0.000539	0.000033	$< 10^{-6}$
LUBM300/regexes/q4_3_0	0.658921	0.002423	0.000049	$< 10^{-6}$
LUBM300/regexes/q4_4_0	0.680183	0.004242	0.000053	$< 10^{-6}$
LUBM300/regexes/q4_5_0	0.682706	0.004196	0.000063	$< 10^{-6}$
LUBM300/regexes/q5_0	0.755279	0.00213	0.000041	$< 10^{-6}$
LUBM300/regexes/q6_0	0.921146	0.000348	0.000034	$< 10^{-6}$
LUBM300/regexes/q7_0	0.578389	0.000128	0.000046	$< 10^{-6}$
LUBM300/regexes/q8_0	0.580071	0.000108	0.00002	$< 10^{-6}$
LUBM300/regexes/q9_2_0	0.612161	0.000121	0.000029	$< 10^{-6}$
LUBM300/regexes/q9_3_0	0.6347	0.000671	0.00004	$< 10^{-6}$
LUBM300/regexes/q9_4_0	0.61649	0.000786	0.000055	$< 10^{-6}$
LUBM300/regexes/q9_5_0	0.612655	0.000321	0.000061	$< 10^{-6}$

2. Сравнение производительности реализаций транзитивного замыкания

Таблица 2: Результаты замеров времени транзитивного замыкания графов, построенных по регулярным выражениям, сек

Регулярное выражение	Кол-во достижимых пар	Транзитивное замыкание через			
		умножение на матрицу смежности		возведение в квадрат	
		Среднее значение	Дисперсия	Среднее значение	Дисперсия
LUBM300/regexes/q1_0 (12)*	1	0.000082	$< 10^{-6}$	0.000062	$< 10^{-6}$
LUBM300/regexes/q10_2_0 ((16 11).(4)*)	2	0.000134	$< 10^{-6}$	0.000087	$< 10^{-6}$
LUBM300/regexes/q10_3_0 ((13 (2 8)).(15)*)	2	0.00016	$< 10^{-6}$	0.000099	$< 10^{-6}$
LUBM300/regexes/q10_4_0 ((11 (16 (9 2))).(1)*)	2	0.00018	$< 10^{-6}$	0.000112	$< 10^{-6}$
LUBM300/regexes/q10_5_0 ((9 (6 (10 (12 1))).(3)*)	2	0.000196	$< 10^{-6}$	0.000122	$< 10^{-6}$
LUBM300/regexes/q11_2_0 (2.1)	3	0.000158	$< 10^{-6}$	0.000116	$< 10^{-6}$
LUBM300/regexes/q11_3_0 (5.(0.11))	6	0.00021	$< 10^{-6}$	0.000168	$< 10^{-6}$
LUBM300/regexes/q11_4_0 (8.(3.(11.1)))	10	0.000275	$< 10^{-6}$	0.000174	$< 10^{-6}$
LUBM300/regexes/q11_5_0 (1.(0.(6.(7.4))))	15	0.00034	$< 10^{-6}$	0.000227	$< 10^{-6}$
LUBM300/regexes/q_12_0 (((10.11))* ((16.12))*)	12	0.00021	$< 10^{-6}$	0.000137	$< 10^{-6}$
LUBM300/regexes/q_13_0 (((14.((16.2))*)* ((8.12))*)	12	0.000214	$< 10^{-6}$	0.000151	$< 10^{-6}$
LUBM300/regexes/q_14_0 (((6.(13.((11.1))*))*).((5 10))*)	17	0.000269	$< 10^{-6}$	0.000196	$< 10^{-6}$
LUBM300/regexes/q_15_0 (((11 13))*.(4 16))*)	3	0.000157	$< 10^{-6}$	0.000099	$< 10^{-6}$
LUBM300/regexes/q_16_0	6	0.000256	$< 10^{-6}$	0.000188	$< 10^{-6}$

(3.(15.(10 (11 5))))					
LUBM300/regexes/q2_0 (7.(4)*)	2	0.00015	$< 10^{-6}$	0.000082	$< 10^{-6}$
LUBM300/regexes/q3_0 (14.((5)*.(16)*)	5	0.000191	$< 10^{-6}$	0.00016	$< 10^{-6}$
LUBM300/regexes/q4_2_0 ((13 2))*	1	0.000095	$< 10^{-6}$	0.000061	$< 10^{-6}$
LUBM300/regexes/q4_3_0 ((0 (4 10))*)	1	0.000112	$< 10^{-6}$	0.000068	$< 10^{-6}$
LUBM300/regexes/q4_4_0 ((1 (11 (3 4))))*	1	0.000133	$< 10^{-6}$	0.000073	$< 10^{-6}$
LUBM300/regexes/q4_5_0 ((1 (5 (15 (8 4)))))*	1	0.000138	$< 10^{-6}$	0.000087	$< 10^{-6}$
LUBM300/regexes/q5_0 (4.((1)*.14))	4	0.000178	$< 10^{-6}$	0.000129	$< 10^{-6}$
LUBM300/regexes/q6_0 ((12)*.(2)*)	3	0.00013	$< 10^{-6}$	0.000078	$< 10^{-6}$
LUBM300/regexes/q7_0 (11.(12.(16)*)	4	0.000178	$< 10^{-6}$	0.000139	$< 10^{-6}$
LUBM300/regexes/q8_0 (12?.(11)*)	2	0.000118	$< 10^{-6}$	0.000077	$< 10^{-6}$
LUBM300/regexes/q9_2_0 ((3 10))*	1	0.000098	$< 10^{-6}$	0.000062	$< 10^{-6}$
LUBM300/regexes/q9_3_0 ((6 (15 3))*)	1	0.000105	$< 10^{-6}$	0.000065	$< 10^{-6}$
LUBM300/regexes/q9_4_0 ((12 (13 (1 10))))*	1	0.000123	$< 10^{-6}$	0.000073	$< 10^{-6}$
LUBM300/regexes/q9_5_0 ((15 (10 (8 (3 11)))))*	1	0.000137	$< 10^{-6}$	0.00008	$< 10^{-6}$

Таблица 3: Результаты замеров времени транзитивного замыкания графов, построенных по тройкам вершина-метка-вершина, сек

Граф	Кол-во дости- жимых пар	Транзитивное замыкание через			
		умножение на матрицу смежности		возведение в квадрат	
		Среднее значение	Дисперсия	Среднее значение	Дисперсия
LUBM300/LUBM300.txt V =358434 E =1449711	6532882	0.473544	0.002369	0.478476	0.000001
LUBM500/LUBM500.txt V =596760 E =2416513	10885048	0.878165	0.004544	0.897825	0.000002
LUBM1M/LUBM1M.txt V =1188340 E =4820728	21833383	1.927764	0.014851	1.995696	0.000004
LUBM1.5M/LUBM1.5M.txt V =1780956 E =7228358	32799959	3.057482	0.030053	3.175128	0.000227
LUBM1.9M/LUBM1.9M.txt V =2308385 E =9369511	42561736	4.049127	0.057947	4.242053	0.000424

3. Выводы

Были произведены замеры производительности реализации пересечения 5-ти крупных графов с 28-ью небольшими, построенными по регулярным выражениям. Результаты показали, что заметный рост времени работы наблюдается и при значительном увеличении числа вершин и рёбер в одном из графов (так, данные в группе больших графов отличаются друг от друга на сотни тысяч вершин и миллионы рёбер), и при выборе более большого графа среди маленьких (построенных по регулярным выражениям), где разница между количествами вершин и рёбер не столь велика.

Сравнение времени работы реализаций транзитивного замыкания показало, что представление с помощью возведения в квадрат заметно быстрее (примерно в 1,5 раза) на маленьких графах, построенных по регулярным выражениям. На больших графах реализация с помощью умножения на матрицу смежности имеет преимущество по времени, и оно тем больше, чем больше обрабатываемый граф, однако в сравнении с общим временем построения матрицы достижимости эта разница не очень велика. Тем не менее на основании этих результатов можно сделать вывод, что при работе с небольшими графами лучше реализовывать транзитивное замыкание через возведение в квадрат, а с большими — через умножение на матрицу смежности. К большим графам точно можно отнести графы, число вершин которых $>3,5$ тысяч, а рёбер $>1,4$ миллионов, но точную границу между "маленькими" и "большими" графами вышеописанными замерами не установить.