Летняя международная XXXI молодежная Школа-конференция по параллельному программированию

ПАРАЛЛЕЛЬНАЯ РЕАЛИЗАЦИЯ ЗАДАЧИ АВТОМАТИЧЕСКОГО ПОСТРОЕНИЯ ЛЕГКОВЕСНЫХ ОНТОЛОГИЙ

Исполнитель проекта: Ертуяк А.Т. 2 курс, кафедра информатики факультета информационных технологий КазНУ им. аль-Фараби
Руководитель проекта: Мансурова М.Е.
Дата доклада: 13.07.2018

План доклада:

1. Постановка задачи
2. Идея решения
3. Реализация
4. Тестирование
5. Заключение

Постановка задачи:

- **Применение анализа формальных понятий.** Использование теории формальных понятий для построения «легковесных» онтологий, в которых онтологии представляются в виде семантической сети.

- **Разделение на кластеры**. Набор данных, представленный в матрице, подвергается кластеризации, в результате которой объекты, которым присущи одинаковые признаки, группируются в один кластер.

Идея решения

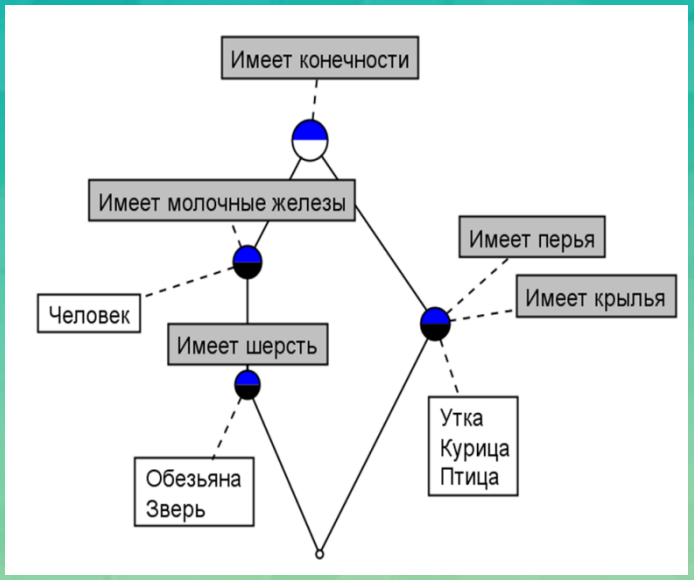
Для разделение на кластеры: алгоритм New Density Based Clustering Algorithm for Binary Data Sets (разновидность алгоритма DBSCAN)

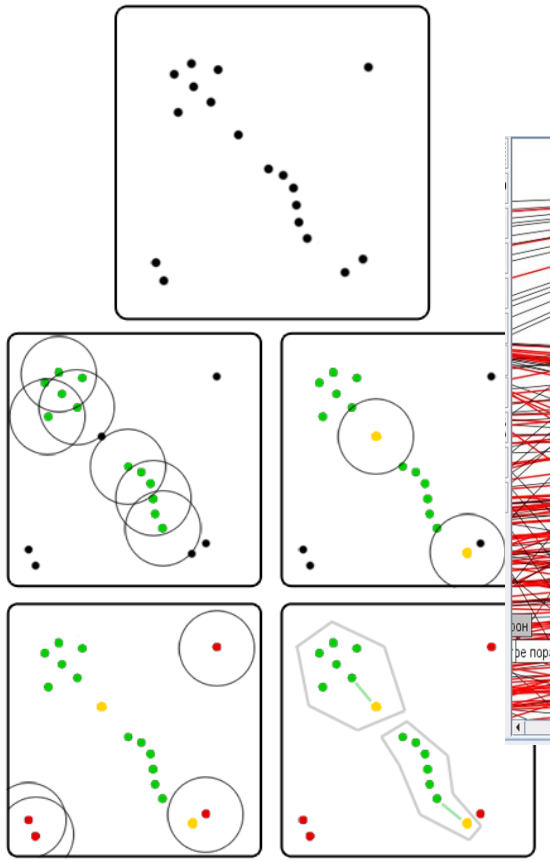
Распараллеливание вычислений: директивы **OpenMP, MPI**

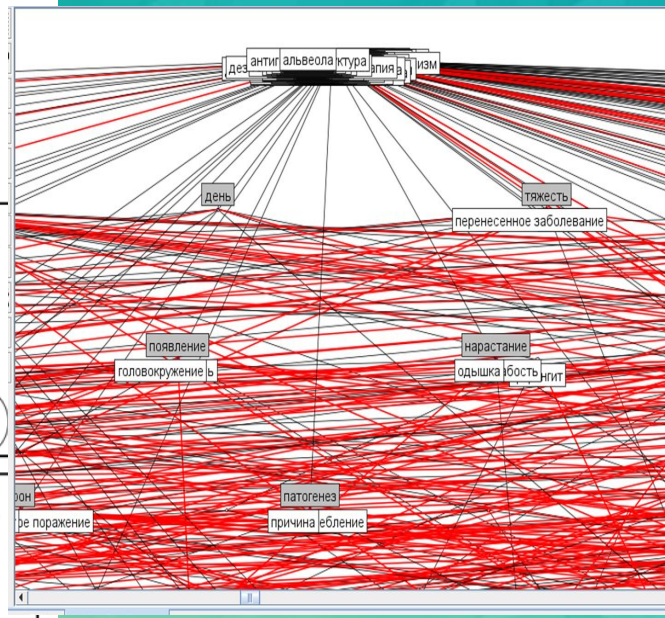
Реализация

1. Изучение алгоритма DBSCAN
2. Построение бинарной матрицы формальных понятий
3. Кластеризаця
4. Ускорение кластеризаций путем директив OpenMP, и MPI на языке C++
5. Сравнение последовательной с параллельной реализацией

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		развитие	случай	поражени	наличие	кровь	вид	болезнь	отсутстви	форма	период	признак	профил
2	абсцесс	1	0	0	0	0	1	0	0	0	0	0	
3	адгезия	0	0	0	0	0	0	0	0	0	0	0	
4	аденовир	0	0	0	0	0	0	0	0	0	0	0	
5	адинамия	0	0	0	0	0	1	0	0	0	0	0	
6	активност	0	0	0	0	0	0	0	0	0	0	0	
7	алкоголь	0	0	0	0	0	0	0	0	0	0	0	
8	аллергич	0	0	0	0	0	0	0	0	0	0	0	
9	альбумин	0	0	0	0	0	0	0	0	0	0	0	
10	альвеола	0	0	0	0	0	0	0	0	0	0	0	
11	амебиаз	0	0	0	0	0	0	0	0	0	0	0	
12	аминотра	0	0	0	0	0	0	0	0	0	0	0	
13	ампицилл	0	0	0	0	0	0	0	0	0	0	0	
14	анализ	0	0	0	0	0	0	0	0	0	0	0	
15	анамнез	0	0	0	0	0	0	0	0	0	0	0	
16	ангина	1	0	0	0	0	0	0	0	0	0	0	
17	анемия	1	0	0	0	0	0	0	0	0	0	1	
18	антибакте	0	0	0	0	0	0	0	0	0	0	0	
19	антибакте	0	0	0	0	0	0	0	0	0	0	0	
20	антибиот	0	0	0	0	0	0	0	0	0	0	0	
21	антибиот	0	0	0	0	0	0	0	1	0	0	0	
								0	0	0	0	0	
								0	0	0	0	0	







Тестирование

Attribute / Sample	1	2	3	4	5	6	7	8	9	10
1	1	1	0	0	0	0	1	1	0	0
2	0	0	0	1	0	0	1	1	1	1
3	1	1	0	0	1	1	0	1	1	1
4	0	0	0	1	1	1	0	1	0	0
5	1	0	0	0	0	0	0	1	1	1
6	1	0	1	1	0	1	0	1	1	0
7	0	1	0	1	0	1	0	0	0	0
8	0	1	0	1	0	1	1	1	1	0
9	1	1	0	1	0	0	1	0	1	0
10	0	1	1	1	1	1	0	0	1	0
11	1	0	1	0	1	0	0	1	0	0
12	1	0	0	1	0	1	0	1	1	0
13	1	1	0	1	0	1	1	0	1	1
14	1	1	1	0	1	0	1	1	0	0
15	0	0	0	1	1	1	0	0	1	1
16	0	1	1	1	1	1	1	1	0	0
17	1	1	1	1	0	1	1	0	0	0
18	1	0	1	0	1	0	0	0	0	0
19	0	1	1	1	0	1	1	0	0	1
20	1	0	0	1	1	0	1	1	0	1
21	0	0	0	1	1	1	0	1	1	0
22	0	0	1	0	0	1	0	1	1	1
23	1	1	0	1	0	1	0	1	0	1
24	0	0	1	0	1	1	0	1	0	1
25	0	1	1	1	1	1	0	0	0	0
26	0	1	1	0	1	0	1	0	1	0
27	0	1	0	1	0	0	0	1	1	0
28	1	1	0	0	0	1	0	0	1	0
29	0	0	0	1	0	0	1	0	1	0
30	0	0	0	0	0	1	1	0	0	0
31	0	0	0	0	0	1	0	1	0	1
32	1	1	1	0	0	0	0	1	1	0
33	0	0	0	0	1	0	0	0	1	0
34	1	1	1	1	1	0	1	1	1	0
35	0	1	0	1	1	0	1	1	1	1
36	0	0	0	0	1	1	0	1	0	1
37	1	1	0	1	0	0	1	1	0	1
38	1	0	0	1	1	1	0	0	1	0
39	0	0	1	1	1	1	1	0	0	0
40	1	1	0	1	1	1	0	1	0	0
41	1	1	0	1	0	0	1	1	0	0
42	0	0	0	0	1	0	1	1	1	0
43	1	1	1	0	1	0	1	1	0	1
44	0	1	1	1	1	0	1	0	0	0
45	1	0	1	1	1	1	1	1	1	1
46	1	0	1	0	0	0	0	0	0	1
47	1	1	0	0	0	1	1	0	0	1
48	0	1	0	0	0	1	0	1	1	1
49	0	0	1	0	1	1	1	1	1	0
50	0	1	1	0	0	1	1	1	0	0
51	0	0	0	1	0	1	1	1	0	0
52	1	0	0	1	1	1	0	1	0	1
53	0	0	0	0	1	0	0	1	0	1
54	1	0	1	0	0	1	0	1	1	1
55	0	0	1	1	0	1	1	1	0	1
56	1	0	1	0	1	0	0	1	1	1
57	0	1	1	1	0	0	0	1	0	1
58	1	0	0	1	1	1	1	1	0	1
59	1	0	1	1	1	1	1	0	1	1
60	1	1	1	0	1	1	0	1	1	1
61	0	1	0	1	0	1	0	0	1	1
62	1	0	1	1	0	1	0	0	0	0
63	1	0	1	1	0	1	0	0	1	1
64	1	0	1	0	1	0	0	0	1	1
65	1	1	0	0	1	0	0	0	1	0
66	1	0	1	0	0	0	1	0	1	1
67	0	0	1	0	0	0	0	1	1	0
68	0	1	0	1	1	0	0	0	0	1
69	0	1	1	1	0	1	0	0	1	0
70	1	1	0	0	0	0	0	1	0	0
71	1	0	1	1	0	0	1	0	1	1
72	0	0	1	0	0	1	0	1	0	1
73	1	0	0	1	1	0	1	1	1	0
74	1	0	0	1	0	1	0	1	1	1
75	0	0	1	0	1	1	0	0	0	0
76	1	1	1	1	0	1	1	0	0	1
77	1	1	0	1	1	0	0	0	0	0
78	0	1	0	1	0	1	1	0	1	0
79	1	0	1	0	1	0	1	1	1	1
80	1	1	1	0	0	1	0	1	0	0
81	1	0	1	0	1	0	1	0	1	1
82	1	0	1	0	1	1	1	1	1	1
83	1	1	1	0	0	0	1	0	0	0
84	1	1	0	0	1	1	0	0	1	1
85	0	0	0	1	0	1	0	0	0	0
86	1	0	1	0	0	1	1	1	0	1
87	0	1	1	0	0	0	0	0	0	0
88	0	0	1	1	1	0	0	1	0	1
89	1	1	1	0	0	1	0	0	0	0
90	1	1	0	1	0	1	0	0	1	1
91	0	0	0	0	1	0	1	1	0	1
92	0	0	1	1	0	1	1	1	1	1
93	1	0	0	1	0	0	0	0	0	1
94	0	0	0	0	0	1	1	1	1	1
95	0	1	1	1	1	0	0	1	1	0
96	1	0	1	1	1	1	1	0	0	1
97	0	1	1	0	1	0	1	1	0	0
98	0	0	1	0	1	1	1	1	0	0
99	0	0	1	0	0	0	1	1	0	1
100	1	1	0	0	0	1	1	1	0	1

```
1       0 40 69
2       75 16 18
3       1 28 34 91 93
4       76 39 64 67
5       25 43 80 96
6       77 7
7       3 20
8       26 7 94
9       78 55
10      27 83
11      4 53 55 73
12      5 11
13      79 31 49 88
14      29 50 84
15      6 84
16      81 48
17      30 35 71
18      82 13
19      8 12
20      32 41
21      51 57
22      9 24 68
23      33 72
24      52 90
25      10 17
26      54
27      56 86
28      14 37 60
29      85 54 98 99
30      36
31      15 38 97
32      87
        58 44 95
        89
        74 95       41      21
        59 2        42      63
        92 45       43      46
        19          44      65 70
        61 62       45      22
        42          46      66
        21          47      23
42      63          48      47
```

Заключение

	100x10	100x100	1000x1000
До OpenMP:	5,025	6,656	45,075
После OpenMP:	0,512	0,516	30,565

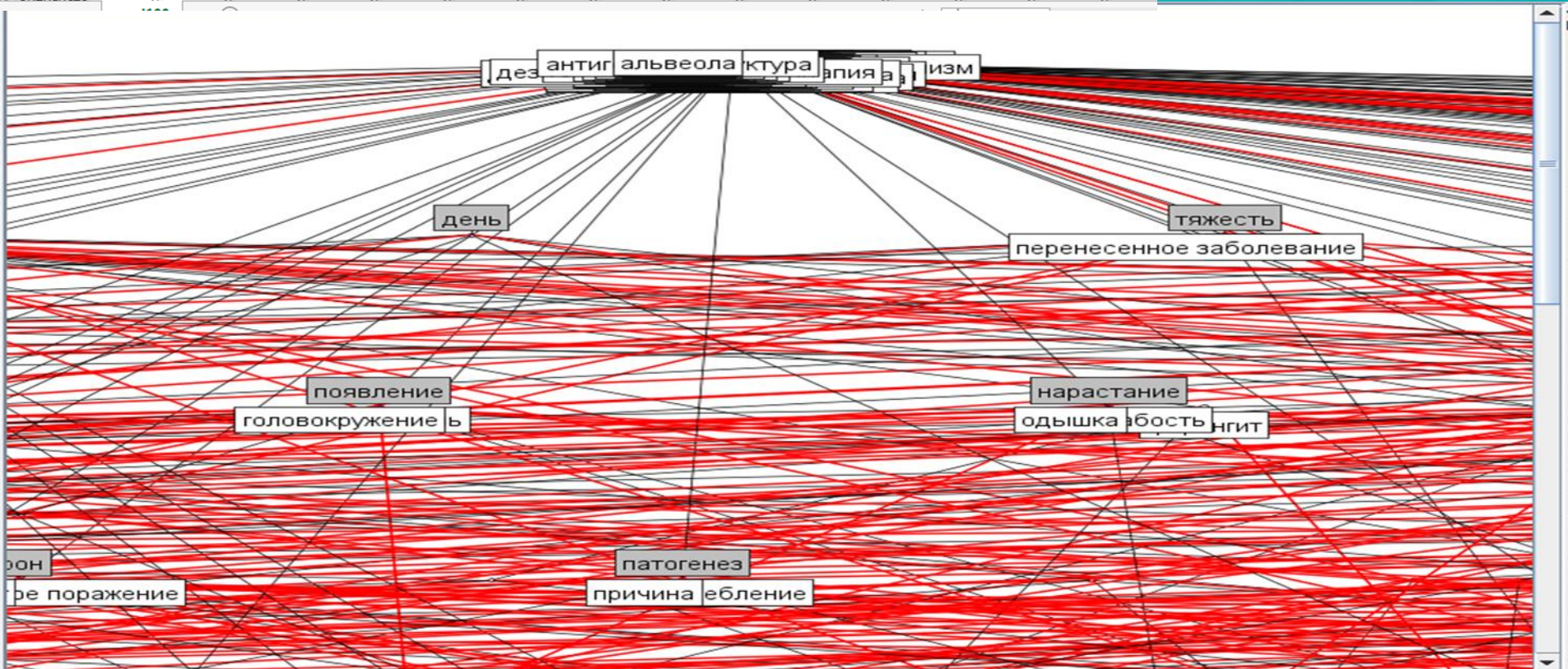
Направление развития проекта:

Ускорить процесс кластеризаций:

- Улучшение эффективности расспараллевания задач

Использование данных с большим объемом:

- Корпус казахского языка

ВСЕМ СПАСИБО!

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		развитие	случай	поражени	наличие	кровь	вид	болезнь	отсутстви	форма	период	признак	профилак	появлени	время	зараж
2	абсцесс	1	0	0	0	0	1	0	0	0	0	0	0	0	0	
3	адгезия	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	аденовир	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	адинамия	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
6	активност	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	алкоголь	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	аллергич	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	альбумин	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	альвеола	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	амебиаз	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
12	аминотра	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
13	ампицилл	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14	анализ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
15	анамнез	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
16	ангина	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
17	анемия	1	0	0	0	0	0	0	0	0	0	1	0	0	0	
18	антибакте	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	антибакте	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
20	антибиоти	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
21	антибиоти	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
22	антиген	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	антигенна	0	0	0	0	0	0	0	0	0	0	0	0	0	0