



3. Недостатки дополненной реальности [Электронный ресурс]
<https://www.mate-expo.ru/ru/article/nedostatki-dopolnennoy-realnosti>

4. Вся статистика интернета на 2019 год – в мире и в России [Электронный ресурс] <https://www.web-canape.ru/business/vsya-statistika-interneta-na-2019-god-v-mire-i-v-rossii/>

Л.А. Гафарова, З.Ф. Камальдинова

ИНФОРМАЦИОННАЯ СИСТЕМА КЛАСТЕРИЗАЦИИ ЧЛЕНОВ ГУБЕРНАТОРСКОГО РЕЕСТРА ПО ДАННЫМ МОНИТОРИНГА ИХ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ ДЕЯТЕЛЬНОСТИ

(Самарский государственный технический университет)

В Самарской области, начиная с 2015 года, формируется Единая Самарская областная система мер по выявлению и развитию творчески одаренной молодежи в сфере науки, техники и технологий и инновационному развитию Самарской области [1]. ЕСМ включает в себя такие научно-образовательные программы как: «Взлёт», «Полёт» и «Орбита» [2]. Все они нацелены на пополнение реестра молодых ученых Самарской области. Для выявления и поддержки творчески одаренной молодежи проводится ежегодный конкурс научно-исследовательских работ. Участники показавшие наиболее высокие результаты попадают в Губернаторский реестр Самарской области.

Для поддержки и повышения творческого и научного потенциала участников необходимо вводить дополнительные научно-образовательные мероприятия, которые должны положительным образом сказаться на уровне и заинтересованности участников. В настоящий момент в НОП «Взлёт» тридцать одна секция [3]. В каждой секции уровень участников различен, в связи с этим подбор эффективных мероприятий становится достаточно затруднительным.

В системе каждый участник имеет следующие характеристики своей научно-исследовательской деятельности: результат рецензии, рейтинги систематичности, активности и достижений [4]. Было выявлено что рейтинги способны влиять на результат рецензии участников (таблицы 1, 2). Данный фактор показал, что для большего эффекта необходимо подбирать мероприятия по характеристикам участников. Для решения поставленной задачи, было решено прибегнуть к кластерному анализу.

Таблица 1 – Градация силы связи

Значение	Интерпретация	Цвет
до 0,2	Очень слабая корреляция	X
до 0,5	Слабая корреляция	X



до 0,7	Средняя корреляция	X
до 0,9	Высокая корреляция	X
свыше 0,9	Очень высокая корреляция	X

Таблица 2 – Исследование «Зависимость результата рецензии от рейтингов систематичности и достижений»

№	Секция	Кол-во участников	Ср. рез-т рецензии + рейтинг систематичности	Ср. рез-т рецензии + рейтинг достижений
1	Авиационная и космическая техника, механика и астрономия	22	0,68	0,46
2	Психология и педагогика	40	-0,10	0,39
3	Техническое творчество, энергетика и транспорт	30	0,48	0,48
4	Большие данные, искусственный интеллект, финансовые технологии и кибербезопасность	14	-0,29	0,42
5	Математика	83	-0,20	0,20
6	Физика	60	-0,08	0,40
7	Генетика, персонализированная и прогностическая медицина	9	0,01	0,01
8	Обществознание и право	23	0,26	0,36
9	Экология и сельское хозяйство	64	-0,05	0,05
10	Биология и медицина	76	0,00	0,26
11	Экономика и менеджмент	27	0,23	0,17
12	Космические технологии	9	-0,21	0,23
13	История	65	0,29	0,27
14	Иностранные языки	60	0,09	-0,02
15	Химия	59	-0,22	0,09
16	Агропромышленные и биотехнологии	25	0,08	0,49
17	Литература	55	-0,22	0,38
18	География и краеведение	54	0,08	-0,03
19	Русский язык	54	0,10	0,47
20	Культурология и искусствоведение	24	-0,02	0,54
21	Нанотехнологии	10	-0,59	-0,49



22	Строительство, архитектура и охрана среды	13	-0,03	-0,23
23	Умный город	12	0,20	0,24
24	Социология	32	0,13	0,09
25	Машиностроение, приборостроение и робототехника	13	0,18	0,34
26	Информатика и информационные технологии	38	0,24	0,46
27	Новые материалы	5	-0,51	0,63
28	Нейросети и природо-подобные технологии	4	-0,44	0,42
29	Беспилотный транспорт и логистические системы	5	0,61	0,51
30	Когнитивные исследования	8	-0,06	-0,10
31	Современная энергетика	9	-0,08	0,18
Общее количество участников 2019 год		1002	0,02	0,24

Перед тем как приступить к кластерному анализу необходимо определиться с методом и метрикой [5, 6]. Сравнительный анализ методов (агломеративный, дивизивный, EM-алгоритм и k-средних) показал, что наиболее подходящим для решения поставленной задачи является метод k-средних. Так как иерархические методы во многом уступают итерационным. А EM-алгоритм продолжается до тех пор, пока во всех кластерах не будет равное количество объектов.

Так же было поведено исследование, которое показало, что выбор метрики способен повлиять на результат рецензии. Результаты продемонстрированы на рисунке 1.

Для проведения кластерного анализа было решено использовать метод k-средних с применением Манхэттенской метрики.

По итогам проведенной работы можно сказать, что разработанная ИС дополнит основной функционал системы, куда предполагается внедрить данный продукт. Проведенное исследование показало, что оказать влияние на научно-исследовательский уровень участников возможно. Путем проведения различных научно-образовательных мероприятий и т.д. Которые помогут замотивировать, наградить и поддержать участников Губернаторского реестра, для пополнения научного реестра молодых ученых Самарской области.

На основе уже имеющихся данных, с помощью разработанной ИС появится возможность формирования индивидуальных групп, которые наилучшим образом позволят распределить участников на группы. Благодаря чему подбор мероприятий окажется более точным, а их влияние эффективным.



	Метод k-средних + Манхэттенская метрика (расстояние городских кварталов)				Метод k-средних + Евклидово расстояние				Метод k-средних + Квадрат Евклидова расстояния			
	ПЕРВАЯ ИТЕРАЦИЯ											
Кластер	1	2	3	4	1	2	3	4	1	2	3	4
Кол-во чел.	3	17	24	5	3	17	24	5	3	17	24	5
	ВТОРАЯ ИТЕРАЦИЯ											
Кластер	1	2	3	4	1	2	3	4	1	2	3	4
Кол-во чел.	9	14	20	6	7	14	12	16	7	14	22	6
	ТРЕТЬЯ ИТЕРАЦИЯ											
Кластер	1	2	3	4	1	2	3	4	1	2	3	4
Кол-во чел.	11	12	20	6	8	15	14	12	8	15	20	6
	ЧЕТВЕРТАЯ ИТЕРАЦИЯ											
Кластер	1	2	3	4	1	2	3	4	1	2	3	4
Кол-во чел.	12	6	20	11	3	17	24	5	8	15	20	6
	ПЯТАЯ ИТЕРАЦИЯ											
Кластер	1	2	3	4	1	2	3	4	1	2	3	4
Кол-во чел.	12	6	20	11	9	14	20	6	8	15	20	6
	ШЕСТАЯ ИТЕРАЦИЯ											
Кластер	1	2	3	4	1	2	3	4	1	2	3	4
Кол-во чел.	12	6	20	11	11	12	20	6	8	15	20	6
	СЕДЬМАЯ ИТЕРАЦИЯ											
Кластер	1	2	3	4	1	2	3	4	1	2	3	4
Кол-во чел.	12	6	20	11	11	12	20	6	8	15	20	6

Рисунок 1 – Исследование «Влияние выбора метрики на результат кластерного анализа»

Литература

1. Камальдинова З.Ф., Куликова Н.В. Анализ итогов проведения областного конкурса исследовательских проектов школьников с применением современных инфокоммуникационных технологий // Вестник Самарского государственного технического университета. Серия: Психолого-педагогические науки, №1(21), 2019. С. 98-110.
2. Единая Самарская областная система мер по выявлению и развитию творчески одаренной молодежи в сфере науки, техники и технологий и инновационному развитию Самарской области (система «Творческая молодежь») [Электронный ресурс]: URL: <http://www.creative-youth.ru> (дата обращения: 28.03.2019).
3. Приказ министерства образования и науки Самарской области об утверждении Порядка включения детей и молодежи в Губернаторский реестр творчески одаренной молодежи Самарской области в сфере науки и техники №244-од от 09.07.2018.
4. Пиявский С.А., Камальдинова З.Ф., Елунин М.Н. О формировании экспериментального сегмента национальной развивающей коммуникационной научно-образовательной среды дополнительного образования творчески одаренных детей и молодежи. /В сборнике: Перспективные информационные технологии (ПИТ 2018) Труды Международной научно-технической конференции. Под редакцией С.А. Прохорова. 2018. С. 1292-1295.
5. Меры расстояния и сходства между объектами [Электронный ресурс]: URL: <https://studfiles.net/preview/1582407/page:3/> (дата обращения: 01.04.2019).
6. Методы кластерного анализа. Иерархические методы [Электронный ресурс]: URL: https://www.intuit.ru/studies/professional_skill_improvements/1210/courses/6/lecture/182?page=2 (дата обращения: 01.04.2019).