

Интеллектуальный анализ данных при категоризации преподавателей вуза на основе наукометрических показателей

О.А. Зятева, Е.А. Питухин, И.В. Пешкова, И.М. Шабалина

Петрозаводский государственный университет

Аннотация: Методами Data Mining был проведен анализ публикационной активности профессорско-преподавательского состава вуза на примере Петрозаводского государственного университета (далее ПетрГУ). Для выявления групп сотрудников с похожими показателями научной деятельности была проведена их кластеризация. В результате преподавательский состав был разделен на восемь кластеров, три из которых включают в себя сотрудников, представляющих настоящее и будущее науки ПетрГУ, а все остальные должны стремиться попасть в эти группы. Представленные результаты статистической обработки показателей могут быть полезны вузу при проведении самоанализа. Руководство вуза может сделать вывод о текущем состоянии научной деятельности, как по отдельному сотруднику, так и по организации в целом. Это позволит ему принять научно-обоснованные управленческие решения в целях улучшения показателей научной деятельности организации.

Ключевые слова: эффективность вуза, научная деятельность, интеллектуальный анализ данных, кластеризация, наукометрические показатели, индекс Хирша, РИНЦ.

Анализ публикационной активности профессорско-преподавательского состава ПетрГУ

В настоящее время одним из важных критериев оценки эффективности университета в целом являются показатели научной деятельности, которые используются при проведении мониторингов, при построении различных рейтингов, учитываются при выделении грантов и т.п. [1–3]. Наиболее значимым наукометрическим показателем, который используется руководством вуза для количественной оценки научной продуктивности ученого, а также при принятии кадровых решений при получении научных званий и др. является индекс Хирша (далее ИХ). На основе этого показателя проводится оценка деятельности профессорско-преподавательского состава (ППС) ряда российских и зарубежных вузов [4-5]. В связи с этим, был проведен анализ публикационной активности за 2016 год в разрезе основных

институтов ПетрГУ и ученых степеней сотрудников на примере индекса Хирша. В качестве информационной базы исследования были собраны и обработаны данные по основным показателям, характеризующим научную деятельность сотрудников вуза, а именно, показатели их публикационной активности и распределение числа их публикаций по годам за период 2000 – 2016 г., имеющих в библиографической базе данных российского индекса научного цитирования (далее РИНЦ) [6]. В таблице № 1 представлено распределение общего числа сотрудников ПетрГУ в разрезе ученых степеней, а также результаты группировки сотрудников институтов по значению индекса Хирша.

Таблица № 1

Распределение образовательных институтов по значению ИХ сотрудников

Подразделе- ние*	Число сотрудников			Число сотрудников по индексу Хирша					Среднее значение индекса Хирша
	Всего	Кандидаты наук	Доктора наук	0	1 - 2	3 - 5	6 - 10	>10	
ИЛГиСН	91	67	17	21	31	18	14	7	3,9
ФТИ	66	45	11	13	25	19	5	4	3,0
ИБЭАТ	74	50	18	21	26	20	7	0	2,4
МИ	106	63	27	34	35	28	7	2	2,2
ИМиТ	71	42	8	19	31	16	5	0	2,0
ИЭП	53	33	7	32	12	5	2	2	1,7
ИПиСН	68	45	9	19	35	13	1	0	1,5
ИПП	69	44	3	29	36	4	0	0	0,9
ИФКиТ	55	17	2	34	13	7	1	0	0,9
ФИ	80	41	8	47	28	5	0	0	0,8
ИИЯ	65	23	2	46	14	5	0	0	0,5
ПетрГУ	798	470	112	315	286	140	42	15	1,9

*Институт биологии, экологии и агротехнологий (ИБЭАТ), Институт иностранных языков (ИИЯ), Институт лесных, горных и строительных наук (ИЛГиСН), Институт математики и информационных технологий (ИМиТ), Институт истории, политических и социальных наук (ИПиСН), Институт педагогики и психологии (ИПП), Институт физической культуры, спорта и туризма (ИФКСиТ), Институт экономики и права (ИЭП), Медицинский институт (МИ), Институт филологии (ИФ) и Физико-технический институт (ФТИ)

Самые многочисленные подразделения – медицинский институт и институт лесных, горных и строительных наук. В их состав входит 13,3% и 11,4% от общего числа сотрудников соответственно. Доля ППС в остальных образовательных институтах в пределах от 6,6% до 9,3%. Наибольшее число остепененных (92%) в составе образовательных институтов наблюдается в ИБЭАТ и ИЛГиСН, менее всего в ИИЯ (38%) и ИФКСиТ (35%).

Также в Таблице 1 приведено распределение численности сотрудников институтов со значением индекса Хирша, разбитым на диапазоны. Сортировка институтов произведена в порядке возрастания доли сотрудников с нулевым значением ИХ. Из представленных данных видно, что большинство сотрудников во всех институтах имеет значение ИХ в пределах от 0 до 2, что составляет около 75% от общего числа ППС вуза. Также, около 40% сотрудников ПетрГУ имеют нулевое значение индекса Хирша. В их число входят сотрудники, у которых по данным базы РИНЦ общее число публикаций равно нулю или те, у кого имеется ряд публикаций, но они не цитируются.

В последнем столбце Таблицы 1 – средние значения ИХ у сотрудников ПетрГУ по институтам в порядке убывания. Лидерами являются преподаватели ИЛГиСН – их среднее значение в два раза превышает среднее значение по вузу в целом. Среди отстающих – ИИЯ, возможно, это связано с невысоким показателем сотрудников с ученой степенью среди ППС.

Средние значения индекса Хирша у сотрудников ПетрГУ в разрезе ученых степеней представлены на Рисунке 1. Наиболее высокие значения индекса имеют доктора наук технических направлений (ИЛГИСН, ФТИ), наименьшие – гуманитарные направления (ФИ, ИПИСН). Стоит отметить, что среди докторов наук и сотрудников, не имеющих ученой степени, идет большой разброс значений ИХ (коэффициент вариации равен 61% и 76% соответственно), у кандидатов наук он составляет 44%. Таким образом

можно заключить, что если за единицу условной эффективности принять ИХ сотрудников, не имеющих степени, то получается, что научная деятельность кандидатов наук в четыре раза эффективнее, чем у неостепененных, а научная деятельность докторов наук в три раза эффективнее, чем у кандидатов (или в 12 раз эффективнее, чем у неостепененных).

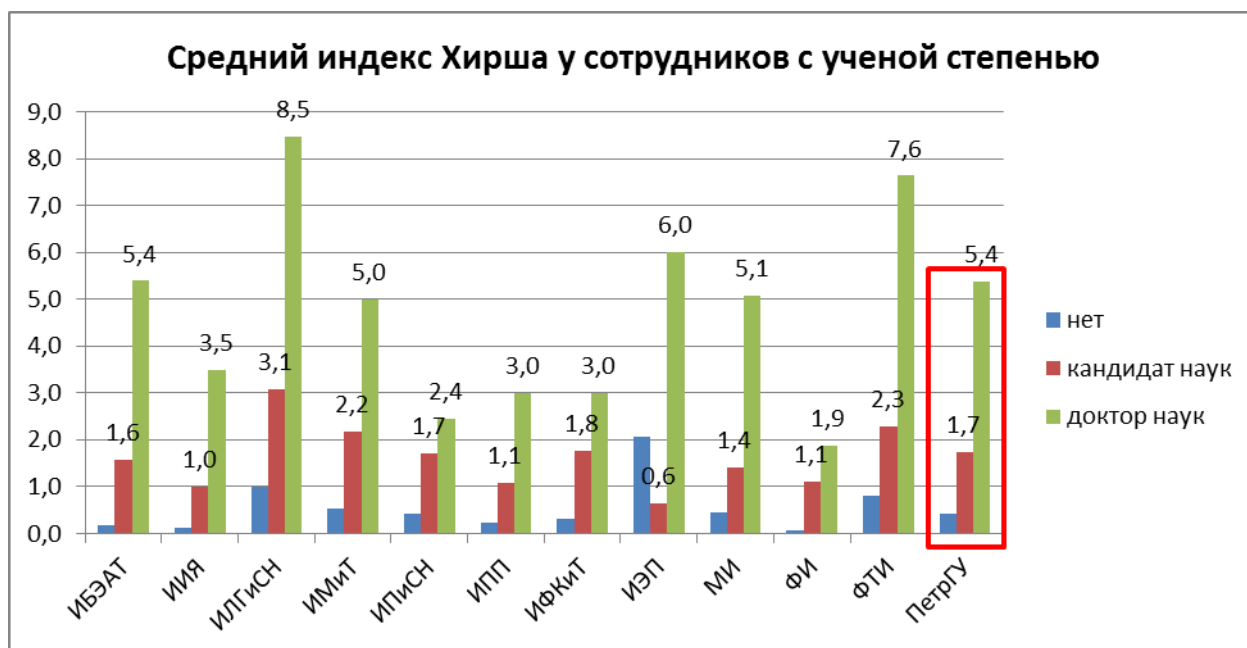


Рис. 1. – Среднее значение индекса Хирша у сотрудников ПетрГУ в разрезе ученых степеней

Рисунок 2 отражает вклад каждого института в общеуниверситетский индекс Хирша ($N=1,9$) в порядке убывания вклада снизу-вверх. Видно, что самый большой вклад вносит ИЛГиСН, наименьший – ИИЯ. Причем ИЛГиСН, МИ и ФТИ суммарно дают больше половины общего значения ПетрГУ.

Существующие сегодня показатели, в частности индекс Хирша, позволяют путем количественного анализа публикаций и их цитируемости сравнивать эффективность деятельности ученых. В результате сравнительного анализа обнаружена большая дифференциация

публикационной активности в разрезе образовательных институтов. Аналогичные различия наблюдаются между сотрудниками без ученой степени и кандидатами и докторами наук [7–8].

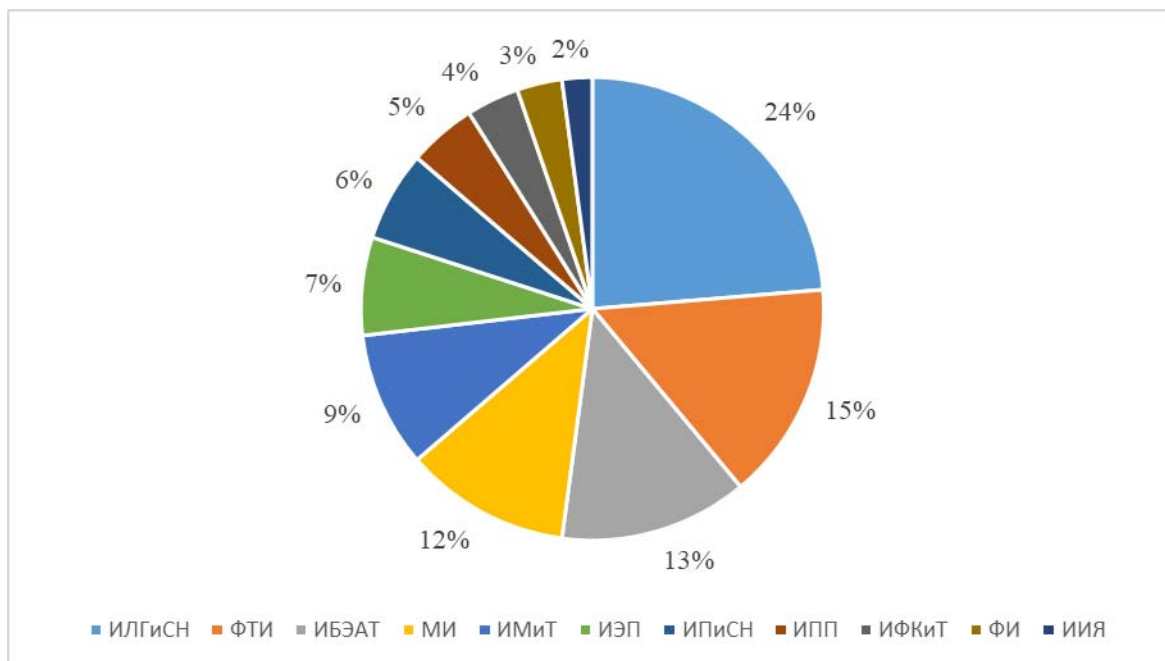


Рис. 2. – Вклад образовательных институтов в общеуниверситетский индекс Хирша

В связи с этим, интересной задачей является разбиение всего множества ППС на устойчивые группы сотрудников со схожими характеристиками публикационной активности.

Кластеризация

Для определения наличия связей между различными показателями (информация о сотрудниках и их наукометрические показатели) были использованы средства интеллектуального анализа данных Excel Add-Ins Analysis Services MS SQL Server. Данные технологии позволяют находить в больших объемах данных скрытые и нетривиальные закономерности, не обнаруживаемые с первого взгляда. В работе использовалось следующее

методы анализа: выделение исключений, анализ ключевых факторов влияния и кластеризация. Входными данными для кластеризации были показатели, полученные в ходе выполнения процедуры анализа ключевых факторов влияния на индекс Хирша. Такими значимыми факторами явились: ученая степень, ученое звание, институт, возраст, стаж, число публикаций автора в РИНЦ, число цитирований публикаций автора в РИНЦ, индекс Хирша и число статей в российских журналах из перечня ВАК. В итоге, все сотрудники были разделены на восемь кластеров. Соответствующие результаты представлены на Рисунке 3.

Переменные	Состояния	Заполнение (все)	Кластер 1	Кластер 2	Кластер 3	Кластер 4	Кластер 6	Кластер 5	Кластер 7	Кластер 8
Размер		798	275	94	90	86	82	79	65	27
Возраст	Среднее	49	48,68	49,84	35,09	53,29	38,94	59,9	59,93	60,76
Возраст	Отклонение	14,26	15,1	8,03	6,05	13,78	6,46	11,35	11,76	12,36
Индекс Хирша	Среднее	2	0	1,43	0,91	5,46	2,74	2,66	0,79	11,98
Индекс Хирша	Отклонение	2,99		0,56	0,59	2,05	1,02	0,95	0,53	7,17
Стаж	Среднее	19	18,79	19,23	9,75	20,45	10,94	29,2	27,84	23,19
Стаж	Отклонение	13,05	13,35	8,81	5,72	15,09	7,14	11,51	12,16	15,92
Ученая степень	кандидат наук	470	46%	94%	47%	44%	90%	62%	80%	12%
Ученая степень	нет	216	53%	5%	53%	0%	10%	0%	15%	7%
Ученая степень	доктор наук	112	2%	2%	0%	56%	0%	38%	5%	81%
Ученое звание	нет	405	73%	31%	97%	21%	67%	1%	24%	9%
Ученое звание	доцент	323	26%	69%	3%	46%	33%	80%	71%	29%
Ученое звание	профессор	70	2%	0%	0%	34%	0%	19%	5%	63%
Факультет _ Институт с 01_01_2018	МИ	106	10%	9%	9%	21%	14%	24%	13%	19%
Факультет _ Институт с 01_01_2018	ИЛГИСН	91	7%	8%	7%	22%	15%	13%	16%	27%
Факультет _ Институт с 01_01_2018	ФИ	74	14%	8%	10%	0%	11%	11%	2%	0%
Факультет _ Институт с 01_01_2018	ИБЗАТ	74	7%	16%	3%	16%	10%	9%	6%	15%
Факультет _ Институт с 01_01_2018	ИПП	69	9%	18%	10%	0%	0%	8%	16%	0%
Факультет _ Институт с 01_01_2018	ИМиТ	69	6%	10%	14%	13%	11%	5%	8%	7%
Факультет _ Институт с 01_01_2018	ИПиСН	68	6%	12%	13%	6%	12%	11%	8%	0%
Факультет _ Институт с 01_01_2018	ФТИ	66	5%	5%	12%	15%	14%	8%	2%	21%
Факультет _ Институт с 01_01_2018
Число публикаций автора в РИНЦ	Среднее	16	0,19	8,81	7,19	44,31	20,08	20,78	5,51	130,19
Число публикаций автора в РИНЦ	Отклонение	32,24	0,65	3,99	4	19,47	8,31	9,59	3,68	98,95
Число статей в российских журналах из перечня ВАК	Среднее	6	0	2,25	2,02	19,21	6,32	7,55	0,77	64,51
Число статей в российских журналах из перечня ВАК	Отклонение	15,43		1,93	1,57	9,65	4,23	4,56	0,86	45,15
Число цитирований публикаций автора в РИНЦ	Среднее	41	0	7,09	2,65	118,57	25,12	28,6	1,85	613,84
Число цитирований публикаций автора в РИНЦ	Отклонение	181,41		4,83	2,75	71,17	17,22	18,81	2,27	759,73

Рис. 3. – Результаты кластеризации сотрудников ПетрГУ

В первый кластер (275 человек) вошли сотрудники, у которых нулевой индекс Хирша, среднее число публикаций автора в РИНЦ равно 0,19, а число цитирований публикаций автора в РИНЦ равно 0. Учёной степени и учёного звания преимущественно нет. При этом средний возраст около 50 лет и половина из них кандидаты наук. Данный кластер состоит из тех, кто снижает среднее значение индекса Хирша в целом по вузу. Если убрать из



базы тех сотрудников, у кого общее число публикаций в РИНЦ равно нулю, то среднее значение ИХ ПетрГУ увеличится в 1,4 раза и станет равным 2,74.

Для второго кластера (94 человека) с таким же средним возрастом сотрудников около 50 лет, но индекс Хирша у которых в среднем 1,43, характерно среднее число публикаций автора в РИНЦ равное 8,81, среднее число цитирований публикаций автора в РИНЦ - 7,09. Качественный состав – преимущественно кандидаты наук с ученым званием доцент. Для среднего значения стажа 20 лет такое число публикаций является незначительным. Скорее всего, данная группа состоит из сотрудников, которые перестали активно публиковаться после защиты кандидатской диссертации.

В третий кластер (90 человек) вошли относительно молодые сотрудники, половина из них кандидаты наук, средний возраст которых около 35 лет и стаж – 10 лет, учёной степени и учёного звания преимущественно нет. Среднее значение числа публикаций автора в РИНЦ составляет 7,19, число цитирований публикаций автора в РИНЦ – 2,02, а ИХ – 0,91. В эту группу вошли малоактивные молодые ученые, которые, вероятно, в ближайшее время не планируют получать ученую степень или звание.

В четвертый кластер (86 человек) вошли сотрудники, имеющие ученую степень и порядка 80% из них – ученое звание, средний возраст которых составляет 53 года. Среднее значение индекса Хирша равно 5,46, числа публикаций автора в РИНЦ – 44,31 и число цитирований публикаций автора в РИНЦ – 118,57. Данные показатели свидетельствуют о том, что эта группа сотрудников активно ведет свою публикационную деятельность, результаты которой интересны другим. Этот кластер – опора университета. Их средний возраст совпадает со средним по вузу, они активно занимаются наукой и могут составить хорошую замену нынешней почетной профессуре.

Пятый кластер (79 человек) состоит из пожилых сотрудников (средний возраст – 60 лет) и стажем 30 лет, большинство из которых имеют ученую степень кандидата или доктора наук, но по званию – доценты. Индекс Хирша в среднем 2,66, число публикаций автора в РИНЦ – 20,78, число цитирований публикаций автора в РИНЦ составляет 28,6. В этот кластер вошли ученые, которые, имея и ученую степень, и ученое звание, продолжают свою публикационную деятельность.

Шестому кластеру (82 человека) характерны средний возраст 39 лет и стаж 11 лет, состоит он преимущественно из кандидатов наук без ученого звания. Индекс Хирша в среднем 2,74, число публикаций автора в РИНЦ – 20,08, число цитирований публикаций автора в РИНЦ составляет 25,12. В данный кластер вошли достаточно молодые сотрудники, которые активно занимаются научной деятельностью, и, возможно, готовятся подавать на ученое звание.

В седьмой кластер (65 человек) вошли сотрудники, у которых средние значения по возрасту и стажу 60 и 28 лет соответственно, в основном кандидаты наук с ученым званием доцент. Среднее число публикаций автора в РИНЦ равно 5,51, а число цитирований публикаций автора в РИНЦ составляет 1,85. В отличие от пятого кластера, среднее значение ИХ равно 0,79. Складывается впечатление, что активную публикационную деятельность они вели перед получением ученого звания, после чего перестали ей заниматься.

Для восьмого кластера (27 человек), самого малочисленного, характерны средний возраст – 61 год, стаж – 23 года. В основном эта группа состоит из докторов наук, имеющих ученое звание. Их индекс Хирша в среднем 11,98, число публикаций автора в РИНЦ – 130,19 и число цитирований публикаций автора в РИНЦ – 613,84. В данный кластер вошли почетные профессора, которые, не смотря на возраст, до сих пор занимаются

наукой и активно публикуют свои результаты. Более молодым сотрудникам стоит брать с них пример.

Полученные восемь кластеров можно упорядочить по среднему возрасту и уровню научной активности сотрудников, связав их в матричную структуру. На рисунке 4 представлена структурная модель в виде трехуровневого графа возможных переходов сотрудников из кластера в кластер, в зависимости от их публикационной активности и с увеличением числа лет. Самый верхний уровень занимают сотрудники, активно ведущие научную деятельность, средний – малоактивные, и, соответственно, нижний – не активные. Стоит отметить, что наиболее вероятные переходы из одного состояния в другое, при условии, что отношение ППС к публикациям останется прежним, отмечены на рисунке 4 жирными стрелками. Такие переходы происходят в рамках одного уровня.

Например, молодые ученые, которые активно публикуются, будут укреплять свои позиции и имеют возможность получить ученые степени или звания (6 → 4) и в дальнейшем стать опорой вуза (4 → 8). Аналогичным образом можно описать процесс для тех, кто не стремится менять свое отношение к публикационной активности. В лучшем случае для них возможен переход 3 → 2 → 5 или 1 → 7. Интересным является то, что, если по каким-либо причинам сотрудник станет уделять больше времени науке и займет активную публикационную позицию, то со временем он может перейти в кластер, на уровень, а то и на два выше, чем был. Соответствующие данной ситуации переходы отмечены на рис. 4 тонкими стрелками. В противоположном случае, сотрудник снижает свои позиции и переходит на уровень ниже, что показано пунктирными стрелками.

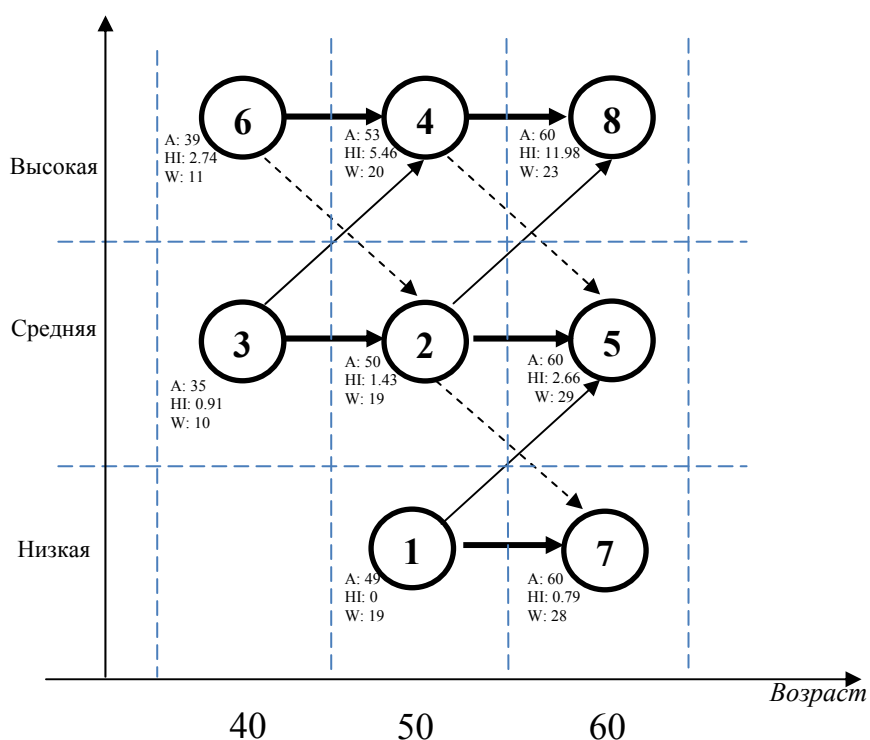


Рис. 4. – Структурная модель переходов преподавателей между выделенными кластерами (А- средний возраст, HI – среднее значение индекса Хирша, W - стаж)

Таким образом, можно выделить три типа переходов, связывающих кластера: жирная стрелка – традиционный сценарий, тонкая стрелка – оптимистичный сценарий, пунктирная стрелка – негативный сценарий.

Предлагаемый инструментарий позволит вузам проводить самоанализ. Результаты проведенного анализа могут быть положены в основу принятия управленческих решений по увеличению значений наукометрических показателей вуза, например, индекса Хирша [9–10].

Руководству вуза следует мотивировать сотрудников из кластеров верхнего уровня оставаться на нем и стимулировать сотрудников нижних уровней перейти на верхний.

Литература

1. Зятева О.А., Пешкова И.В., Питухин Е.А. Разработка системы планирования и контроля научной деятельности вуза // Устойчивое развитие науки и образования. 2017. №2. 168-170.
 2. Zyateva O., Pitukhin E., Peshkova I. Impact of university performance indicators on their position in rankings // EDULEARN16 Proceedings: 8th International Conference on Education and New Learning Technologies. 2016. – pp. 8751-8759.
 3. Котенко Ю.С., Названова И.А., Подопригора М.Г. Проблемы современного вуза и маркетинговые методы их выявления и оценки // Инженерный вестник Дона, 2013, №2 URL: ivdon.ru/ru/magazine/archive/n2y2013/1631.
 4. Пешкова И.В., Васильев А.С. Динамика наукометрических показателей ученых Ухтинского государственного технического университета// X Международная научно-практическая конференция «Образование и наука в современных условиях». Чебоксары, 2017. С. 210–212.
 5. Alfred R., Tanakinjal G.H., Obit J.H. Too few too far: research productivity assessment in Malaysia based on the h-index analysis // INTED2012 Proceedings: 6th International Technology, Education and Development Conference (5-7 March, 2012). – Valencia, Spain, 2012. – pp. 1925–1934.
 6. Зятева О.А., Пешкова И.В., Питухин Е.А. Изменение подхода к оценке наукометрических показателей в РИНЦ: приобретения и потери // Инженерный вестник Дона, 2016, №2 URL: ivdon.ru/magazine/archive/n4y2016/3757/.
 7. Gurtov V.A., Shchegoleva L.V. Comparative analysis of PHDs' publication activity in Russia and other countries // EDULEARN15 Proceedings: 7th International Conference on Education and New Learning Technologies (6–8 July, 2015). – Barcelona, Spain, 2015. – pp. 0985–0989.
-



8. Гуртов В.А, Щеголева Л.В. Соискатель ученой степени доктора наук: публикационная активность // Университетское управление: практика и анализ, 2015, №2(96). С.47-56.
9. Пешкова И.В. О публикационной активности молодых ученых ПетрГУ // Наука, образование, инновации в приграничном регионе. Петрозаводск, 2016. С. 17–18.
10. Бутвило А.И., Коржов С.Т., Кривоноженко А.Ф., Сюнёв В.С. Институты комплексных междисциплинарных исследований как инструмент интенсификации научно-исследовательской деятельности современного вуза // Университетское управление: практика и анализ, 2015, №5(99). С.30-38.

References

1. Zyateva O.A., Pitukhin E.A., Peshkova I.V. Ustojchivoe razvitie nauki i obrazovaniya. 2017. №2. pp. 168-170.
2. Zyateva O., Pitukhin E., Peshkova I. EDULEARN16 Proceedings: 8th International Conference on Education and New Learning Technologies. 2016. pp. 8751-8759.
3. Kotenko Yu.S., Nazvanova I.A., Podoprigora M.G. Inzhenernyj vestnik Dona (Rus), 2013, №2 URL: ivdon.ru/ru/magazine/archive/n2y2013/1631.
4. Peshkova I.V., Vasil'ev A.S. X Mezhdunarodnaja nauchno-prakticheskaja konferencija «Obrazovanie i nauka v sovremennyh uslovijah». Cheboksary. 2017. pp. 210–212.
5. Alfred R., Tanakinjal G.H., Obit J.H. INTED2012 Proceedings: 6th International Technology, Education and Development Conference. 2012. pp. 1925–1934.
6. Zyateva O.A., Pitukhin E.A., Peshkova I.V. Inzhenernyj vestnik Dona (Rus), 2016, №2 URL: ivdon.ru/ru/magazine/archive/n4y2016/3757/.



7. Gurtov V.A., Shchegoleva L.V. EDULEARN15 Proceedings: 7th International Conference on Education and New Learning Technologies. 2015. pp. 0985–0989.

8. Gurtov V.A., Shchegoleva L.V. Soiskatel' uchenoj stepeni doktora nauk: publikacionnaja aktivnost' [Seeker of doctor of sciences (D. Sc.): publication activities]. Universitetskoe upravlenie: praktika i analiz. 2015, №2(96). pp. 47-56.

9. Peshkova I.V. Nauka, obrazovanie, innovacii v prigranichnom regione. Petrozavodsk. 2016. pp. 17–18.

10. Butvilo A.I., Korzhov S.T., Krivonozenko A.F., Syunyov V.S. Instituty kompleksnykh mezhdistsiplinarykh issledovaniy kak instrument intensivifikatsii nauchno-issledovatel'skoy deyatel'nosti sovremennogo vuza [Institutes integrated interdisciplinary research as a tool of intensification of research activities of the modern University]. Universitetskoe upravlenie: praktika i analiz. 2015. №5 (99). pp. 30-38.