

Инициативный академический проект участников группы высокого профессионального потенциала Национального исследовательского университета «Высшая школа экономики»

1. Наименование проекта

Рекомендательная система поиска релевантных статей и грантовых конкурсов на основе персонифицированной информации преподавателя-исследователя "Academic teambuilder".

2. Направленность проекта

Развитие в университете управленческих практик, связанных с накоплением знаний о публикационной активности сотрудников и анализе взаимодействия структурных подразделений и отдельных сотрудников при организации научно-исследовательской работы

3. Детализация направления проекта

Работа исследователя включает формирование академических навыков, установление связей в академической среде и интеграцию результатов научной деятельности в академическом сообществе в виде публикаций в международных индексируемых высоко цитируемых журналах, а также перевод качественных характеристик исследований в материальное обеспечение сопровождения исследований путем участия в различных конкурсах на получение грантов.

Данный проект ставит своей целью решить следующие задачи:

1. Проанализировать структурные зависимости взаимодействия преподавателей различных подразделений, а также уровень вовлеченности подразделений НИУ ВШЭ в совместные проекты на основе графа соавторов для публикаций сотрудников НИУ ВШЭ. Мы ставим целью выявить структурные проблемы низкой публикационной активности отдельных сотрудников и подразделений, и предложить им подборку статей авторов по схожим интересам, с которыми наиболее вероятно появление совместных статей по схожим научным интересам. Будет построена рекомендательная система, реализующая модель «Научной свахи» для начинающих исследователей, позволяющая ускорить процесс их интеграции в сформировавшиеся научные коллективы университета.
2. На основе выделения ключевых слов в публикациях сотрудников и их научных интересов реализовать автоматизированную рассылку персонифицированной подборки предложений по участию в конкурсах на получение грантов и стипендий, наиболее релевантных каждому исследователю. Планируется провести работу по агрегации имеющихся открытых источников с информацией о конкурсах, и на основе автоматической обработки текстов публикаций сотрудников НИУ ВШЭ построить рекомендательную систему, подбирающую список тематически наиболее близких конкурсов для преподавателя. Также возможно использование предложенного метода для рекомендации журналов, наиболее релевантных исследовательскому полю и уровню исследователя.

4. Персональный состав участников проектной группы (список)

№ п/п	ФИО	Должность	Факультет, Научно-исследовательский институт, Научно-учебная лаборатория	Категория кадрового резерва
1.	Карпов Илья Андреевич	Младший научный сотрудник	Международная лаборатория прикладного сетевого анализа	Новые исследователи
2.	Макаров Илья Андреевич	Старший преподаватель, заместитель руководителя департамента	Департамент анализа данных и искусственного интеллекта, Факультет компьютерных наук	Новые преподаватели
3.	Герасимова Ольга Александровна	Стажер-исследователь	Департамент анализа данных и искусственного интеллекта, Международная научно-учебная лаборатория «Интеллектуальных систем и структурного анализа»	Новые исследователи
4.	Мещерякова Наталья Геннадьевна	Стажер-исследователь	Международная Научно-исследовательская Лаборатория Анализа и Выбора Решений	Новые исследователи

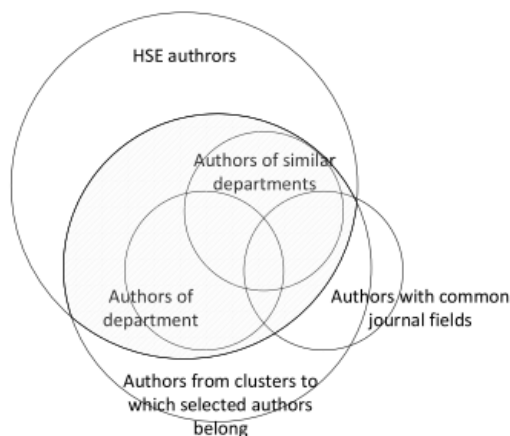
История создания группы
Группа создана по итогам предварительного семинара по обсуждению ИАП-2017. Два инициативных предложения от И.А. Макарова по налаживанию академических связей внутри университета и И.А. Карпова по автоматизации системы доставки контента по грантовым конкурсам и конференциям нашли отражения в создании проектной группы.
Варианты коммуникации группы
Все участники рабочей группы работают в Москве, встречаются лично на Мясницкой, 20

5. Обоснование необходимости проекта (текст объемом не более 2 стр.). *В тексте кратко опишите уже имеющиеся в НИУ ВШЭ наработки по аналогичной тематике и укажите, что нового даст реализация вашего проекта.*

Рекомендательная система для поиска соавторов среди сотрудников НИУ ВШЭ

Современные библиографические базы данных содержат значительное количество информации о публикации деятельности научных сообществ. Исследователи регулярно сталкиваются с проблемой выбора соавтора для проведения совместных исследований или поиска авторов, с чьими работами стоит ознакомиться в первую очередь.

Одной из основных целей проекта является создание рекомендательной системы для нахождения возможного соавтора относительно научных интересов. Рекомендательная система – это программа, которая пытается предсказать, какие объекты будут интересны пользователю, имея определенную информацию о нем.



Система основана на данных о публикациях сотрудников НИУ ВШЭ, и позволяет спрогнозировать появление статьи в соавторстве, а также предполагаемый уровень публикации, измеряемый как квартиль в системах WoS и SJR. Предлагается протестировать эффективность данной системы на участниках программы «Кадровый резерв» и обычных сотрудниках для ускорения интеграции новых исследователей в университетское научное сообщество.

Рекомендательная система для рассылки целевых конкурсов среди сотрудников НИУ ВШЭ

Одним из приоритетов развития подразделений является привлечение внешнего финансирования, в т.ч. через гранты и конкурсы, регулярно публикуемые на

соответственных площадках. Отсутствие эффективного способа актуализации научных интересов и компетенций отдельного исследователя значительно затрудняет доведение информации о предстоящем конкурсе до потенциальных исполнителей. Имеющиеся практики основаны на рассылке информации отдельным исследователям или ответственным представителям подразделений, доводящим информацию до своих коллег посредством "слабых связей". Недостатком такого подхода являются значительные временные затраты на мониторинг и анализ полученной рассылки и административная неэффективность доведения информации до сотрудников. Предлагается создание прототипа информационной системы целевого уведомления сотрудника о предполагаемом конкурсе на основании автоматического текстового и сетевого анализа публикаций, выполняемых исследователем. Данная процедура позволит держать профиль исследователя в актуальном состоянии без человеческого вмешательства и значительно повысить качество предварительной фильтрации. При должном уровне релевантности поступающих материалов представляется возможным исключить промежуточные звенья в описанной ранее цепочке донесения информации и доводить потенциальные проекты до непосредственного исполнителя, что по предположению авторов проекта повысит вовлеченность подразделений во внешние конкурсные процедуры.

Анализ публикационной активности и межструктурного взаимодействия подразделений

Другим аспектом исследования графа публикаций является прогнозирование публикационной активности подразделений и научных направлений в рамках программы развития всего университета в целом. Анализируя описательные статистики подграфов, соответствующих административным подразделениям университета, можно не только спрогнозировать качество и количество публикаций на 1–3 года с точностью более 90%, но и выявить существующие проблемы, которые скрыты за данными, идущими в отчеты. В частности, анализ распределений количества публикаций и степеней вершин позволяет выявить ситуации, когда департамент выполняет показатели публикационной активности за счет отдельных своих представителей, в то время как большая часть преподавателей мало публикуется и ограничена в своем круге коллег. Именно для таких подразделений необходимо дополнительное стимулирование изменить сложившийся консервативный уклад с помощью предложения новых тем для междисциплинарных исследований.

6. Цели проекта (текст объемом не более 0,5 стр.)

Проект имеет две цели:

- 1) улучшение управленческих практик путем создания и пилотного внедрения прототипа информационной системы уведомления сотрудников о внешних конкурсах. В случае успешной реализации, проект позволит укрепить внешние связи с заказывающими организациями
- 2) укрепление внутриуниверситетских связей благодаря рекомендательному механизму подбора команды для совместного выполнения проекта или исследования

7. Укажите целевую аудиторию и кратко опишите механизмы работы с ней (текст объемом не более 0,5 стр.)

Целевая аудитория – научно-педагогический коллектив НИУ ВШЭ, реализующий научно-исследовательскую деятельность, сотрудники дирекции научных исследований. Подразумевается, что в качестве тестирования рекомендательной системы соавторства будут выбраны добровольцы из числа новых преподавателей,

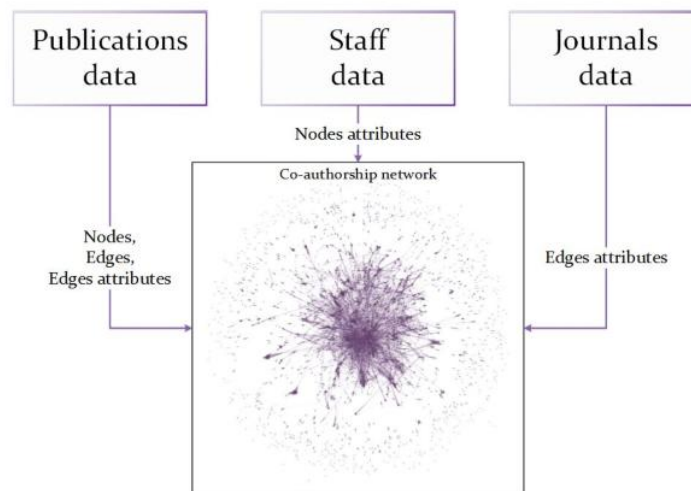
преподавателей с низкой публикационной активностью и а также члены кадрового резерва, взявшие на себя обязательства по публикационной активности.

Описание работы системы и предложение воспользоваться ею, а также организация межличностного взаимодействия будет основано на общении через корпоративную почту, организацию научных семинаров с приглашением потенциальных соавторов, а также brain-storm технологии коворкинга и тимбилдинга, применяемые в таких компаниях, как Google и Facebook для улучшения интеграции новых сотрудников в проектные команды.

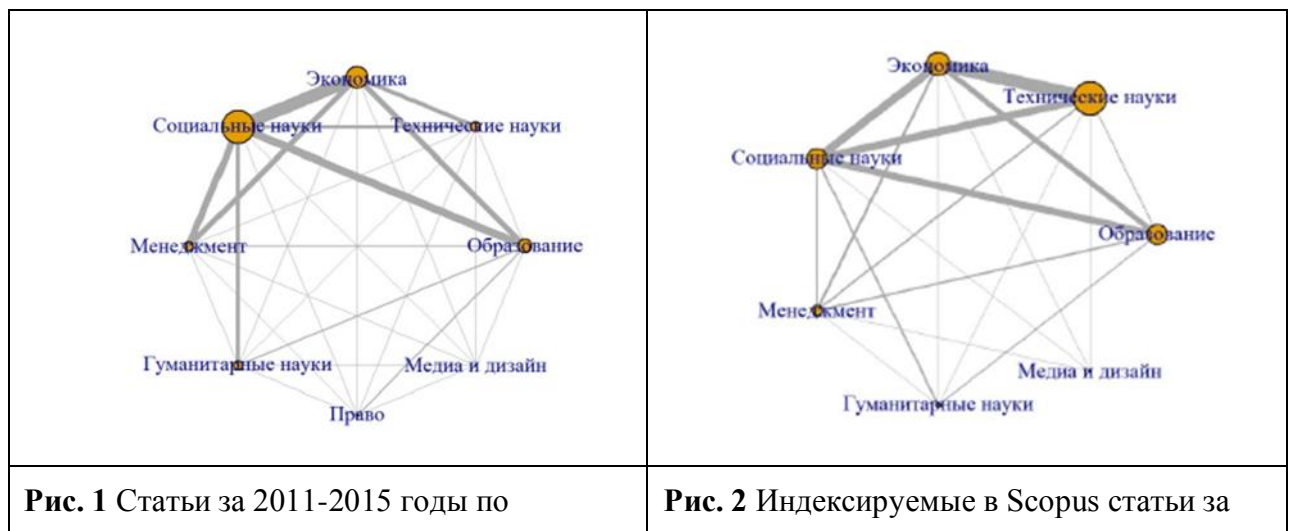
8. Содержание проекта (текст объемом от 0,5 до 1,5 стр.)

Используя данные о публикациях сотрудников НИУ ВШЭ, мы можем построить граф по научным направлениям, вершинами которого будут научные направления, а ребрами – наличие публикаций в соавторстве между авторами, публикующимися в журналах по разным направлениям, что является основой для анализа продвижения НИУ ВШЭ в предметных рейтингах на количественном и качественном уровнях.

Co-authorship network building



Визуализируем граф областей так, что размером вершин будет число публикаций, а толщина ребер будет соответствовать число совместных публикаций. Рассмотрим два графа, соответствующие всем (рис. 1) и только индексируемым в Scopus (рис. 2) публикациям за 2011-2015гг.



На рисунке 1 видно, что все области имеют совместные статьи. Больше всех публикуются вместе экономисты и представители социальных наук. Эта связь кажется достаточно логичной, поскольку политика и социологические исследования тесно связаны с экономикой.

Однако, наиболее интересным фактом, который виден на рисунке 2, является размер вершины «технические науки», она значительно больше остальных. Наиболее тесные соавторства — это связи между областями экономики с техническими науками и социальными науками с техническими науками. Это позволяет сделать вывод, что наиболее качественные статьи должны иметь математический аппарат.

За последние пять лет доля публикаций в Scopus в области экономики увеличилась, что говорит о том, что основной профиль университета стремится увеличить качество научной деятельности с целью увеличения рейтинга университета, однако, на данный момент — это увеличение имеет экстенсивный характер и прямую связь с численностью сотрудников, занимающихся экономическими исследованиями.

	Образование	Технические науки	Экономика	Социальные науки	Менеджмент	Гуманитарные науки	Право	Медиа и дизайн
Число авторов	19,38	19,45	14,15	14,74	12,00	20,24	12,41	11,86
Число публикаций	442,15	454,52	281,65	402,01	187,38	256,88	264,71	146,86
Публикации на число авторов	27,66	26,55	21,30	30,51	15,44	13,67	22,30	10,94
Авторов на число публикаций	0,09	0,06	0,07	0,05	0,09	0,12	0,07	0,10
Публикаций на число авторов	1,71	1,89	1,61	1,76	1,63	1,63	1,56	1,25
Авторов на число публикаций	0,66	0,54	0,65	0,62	0,65	0,71	0,67	0,82
Статей в Scopus	5,08	47,00	4,50	4,39	2,13	2,79	0,24	1,00
Статей в Scopus на число авторов	0,21	2,14	0,36	0,29	0,16	0,16	0,03	0,08
Статей в Scopus за 3 года	3,77	13,00	2,71	2,12	1,41	1,70	0,24	0,57
Статей в Scopus на число авторов за 3 года	0,16	0,59	0,21	0,15	0,12	0,09	0,03	0,07
Статей в Scopus 1 и 2 кварталей	2,31	36,07	2,88	2,00	1,63	2,48	0,00	1,00
Статей в Scopus 1 и 2 кварталей на число авторов	0,12	1,64	0,22	0,13	0,12	0,13	0,00	0,08
Статей в Scopus 1 и 2 кварталей за 3 года	1,69	9,41	1,40	0,91	1,06	1,70	0,00	0,57
Статей в Scopus 1 и 2 кварталей на число авторов за 3 года	0,10	0,40	0,10	0,06	0,10	0,09	0,00	0,07
Среднее число соавторств	12,91	13,10	12,71	15,98	9,29	8,21	14,00	8,89
Компонент связности	9,00	10,14	7,33	6,70	6,31	9,73	5,29	7,37
Размер наибольшей компоненты связности	9,62	7,48	6,65	8,28	5,97	9,91	7,47	4,43
Среднее расстояние	1,51	1,48	1,43	1,55	1,35	1,52	1,42	1,19
Диаметр	6,31	13,21	7,40	7,97	5,31	6,33	6,18	4,57
Коэффициент кластеризации	0,33	0,34	0,36	0,49	0,44	0,46	0,44	0,31
Плотность	0,20	0,12	0,13	0,19	0,16	0,14	0,29	0,09
преподавателей	3,39	3,75	3,23	2,62	1,81	3,13	2,38	2,60
публикаций преподавателей за 3 года	3,80	4,37	2,41	3,70	3,43	4,73	6,17	2,03
Средняя оценка преподавателей	25,71	19,99	16,09	17,91	15,69	18,85	19,06	18,25
старших преподавателей	3,80	4,76	2,98	4,71	5,48	5,18	7,05	2,03
публикаций старших преподавателей за 3 года	25,71	21,81	19,93	22,80	25,10	20,64	21,79	18,25
Средняя оценка старших преподавателей	1,43	1,55	2,71	1,67	0,60	6,29	1,29	0,20
Число доцентов	25,71	21,81	19,93	22,80	25,10	20,64	21,79	18,25
публикаций доцентов за 3 года	1,43	1,55	2,71	1,67	0,60	6,29	1,29	0,20
Средняя оценка доцентов	3,45	4,63	3,44	8,00	0,50	6,46	0,77	6,00
Число профессоров	1,86	2,00	2,41	2,19	1,89	6,86	2,80	1,00
Публикаций профессоров за 3 года	10,81	5,96	4,29	9,72	1,34	6,65	2,24	10,00
Средняя оценка профессоров	25,71	19,52	17,06	21,35	10,56	19,20	15,05	10,00

Таблица 1. Статистика по направлениям исследований в НИУ ВШЭ

Использование машинного обучения и анализа данных при организации и планировании деятельности научных подразделений стало стандартом за рубежом. Создание и тестирование средств анализа публикаций в Российских вузах позволит не только более точно оценить и смоделировать деятельность российских исследователей, но также позволит ввести систему эффективного планирования и автоматического распределения грантовых средств на поддержку публикующихся ученых.

Реализация своевременной рассылки по Call for Papers релевантных конференций также поможет улучшить своевременную подготовку статей на топовые конференции и правильный тайм-менеджмент исследовательской деятельности, поскольку планирование публикации на конференцию Core A* занимает около 1,5 лет до момента проведения самой конференции (дедлайн подачи истекает, обычно, за 6-9 месяцев до конференции).

9. Распределение обязанностей в проектной группе и вовлеченные специалисты

Участники (ФИО)	Роль в проекте
Карпов И.А.	Обработка текстовой информации из базы публикаций портала и построение рекомендательной системы предложений участия в грантовых конкурсах
Макаров И.А.	Анализ графа публикаций на основе соавторства сотрудников НИУ ВШЭ; построение рекомендательной системы соавторов на основе предметных категорий в Scopus и информации о публикациях сотрудника
Герасимова О.А.	Мониторинг публикационной активности подразделений и выработка рекомендаций по установлению сотрудничества между структурными единицами
Мещерякова Н.Г.	Прогнозирование сохранения научного руководства, появления публикаций со студентами и ранжирование преподавательской активности по созданию научных школ на базе НИУ ВШЭ
Все участники	Агрегация информации о грантовых конкурсах и конференциях, и работа с созданием прототипа программного комплекса в виде аналитической программы и/или онлайн-сервиса, в идеале, интегрированного с порталом НИУ ВШЭ
Вовлеченные специалисты, внешние партнеры, преподаватели и научные сотрудники НИУ ВШЭ, др. (если есть в проекте)	
доц. Игнатов Д.И.	Эксперт в области построения рекомендательных систем и машинного обучения и майнинга данных
проф. Жуков Л.Е.	Эксперт в области анализа сетевых структур и методов машинного обучения на данных со сложной структурой
www.portal.hse.ru	Работа с API портала НИУ ВШЭ
Структурное подразделение – партнер проектной группы	
Дирекция научных исследований и разработок	

10. План интеграции проекта в жизнь НИУ ВШЭ (текст объемом не более 1 стр.)

Рекомендательная система, позволяющая спрогнозировать появление статьи в соавторстве и уровень публикации, измеряемый как квартиль в системах WoS и SJR, может стать инструментом кадровой политики относительно определения успешности пребывания в кадровом резерве на основе апробации указанной методики, а также предполагаемый уровень публикации.

Внедрение системы рекомендации конкурсов предполагается выполнить путем выполнения рассылки о проводимых конкурсах среди добровольцев в выбранной области знаний.

Предлагается протестировать эффективность системы на участниках программы «Кадровый резерв» и обычных сотрудниках для ускорения интеграции новых исследователей в университетское научное сообщество.

11. Ожидаемые результаты (текст объемом не более 1 стр.)

Итоги проекта можно разделить на несколько составляющих:

Первое: Предварительно получен результат для графа публикаций, на котором представлены соавторы 2-4 статей разных областей исследования, - оказывается, что среднее число человек необходимое для того, чтобы связаться с любым другим человеком через научное сотрудничество характеризует число основных направлений исследований в университете. Сейчас это число колеблется от 7.5 до 9, определяя основные направления исследования в соответствии с реструктуризацией факультетов. Мы ожидаем, что внедрение системы рекомендации соавторства позволит уплотнить граф публикаций в соавторстве с сотрудниками НИУ ВШЭ и упрочить межструктурные взаимодействия, в том числе, на уровне САЕ.

Второе: Информирование сотрудников НИУ ВШЭ о конкурсных заявках на гранты и объявления о начале приема статей на топовые конференции, индексируемые в Скопус. Проект позволит упростить процедуру донесения информации о конкурсах до непосредственных исполнителей без временных задержек и посредников, что, по предположению авторов проекта, повысит вовлеченность подразделений во внешние конкурсные процедуры и укрепит внешние связи с заказывающими организациями. Грамотная рассылка информации по релевантным конференциям поможет самоорганизовать научную деятельность сотрудника, заменив приходящий на корпоративные адреса спам по predatory journals and conferences адекватной подборкой конференций, подходящих под профиль исследователя.

Третье: Полученная в ходе работы комплексная модель исследователя может быть в дальнейшем использована для оценки эффективности структурных подразделений и задач управления интеллектуальным капиталом НИУ ВШЭ, повышая конкурентоспособность индивидуальных преподавателей и структурных подразделений в российском и международном научном пространстве.

12. План проекта

№ п/п	Мероприятие, ход реализации	Описание	Эффекты / результаты	Сроки
1	Этап 1. Разработка текстовой и сетевой модели исследователя («01»апреля 2017 – «31» июля 2017 г.)			
1.1	Текстовый анализ статей сотрудников, построение дистрибутивной	Разработка процедуры текстового анализа статей	База данных структурированной информации, описание	«01»апреля 2017 – «31» июня 2017 г

	модели	на русском и английском языке	предложенного метода автоматической обработки	
1.2	Построение графа связей исследователя на основе наукометрической информации и результатов текстового анализа	Построение векторного представления профиля исследователя для последующего сетевого анализа	База данных структурированной информации, описание предложенного метода автоматической обработки	«01»апреля 2017 – «31» июня 2017 г
2	Этап 2. Разработка прототипа рекомендательной системы («01» августа 2017 – «30» ноября 2017 г)			
2.1	Разработка метода предсказания соавторов	Алгоритм рекомендации научных контактов среди сотрудников НИУ ВШЭ на основании предложенного векторного представления	описание предложенного метода автоматической обработки	«01» августа 2017 – «30» сентября 2017 г
2.2	Разработка метода ранжирования уведомлений о конкурсах и конференциях	Алгоритм рассылки в соответствии с моделью исследователя соавторов	описание предложенного метода автоматической обработки	«01» августа 2017 – «30» сентября 2017 г

		среди сотрудников НИУ ВШЭ		
2.3.	Разработка прототипа программных средств рекомендательной системы	Программная реализация предложенных методов	Сервис почтовой рассылки, веб сайт с визуализацией сетевых связей сотрудников	«01» октября 2017 – «30» ноября 2017 г
2.4.	Написание отчета и документации по проекту	Написание отчетной документации по проекту	Отчет о проекте	«01» августа 2017 – «30» ноября 2017 г

13. Способы информирования о проекте

Представление на мероприятиях кадрового резерва, почтовая рассылка, онлайн-сервис (опционально).

«9» марта 2017 г.

Подпись руководителя
организационной группы

(Карпов И.А., м.н.с.
международной
лаборатории прикладного
сетевое анализа)

«9» марта 2017 г.

Подпись руководителя
организационной группы

(Макарова И.А., зам. рук.
департамента анализа
данных и искусственного
интеллекта)