Organisation Europeenne Pour La Recherche Nucleaire

GENEVE, SUISSE
GENEVA, SWITZERLAND

N 2020-13

Mail address: PH Department, CERN,
CH-1211 GENEVE 23, Switzerland

Telex / Telex : 419000 CER CH
Télégramme/Telegram : CERNLAB-GENEVE
Téléfax/fax : +41 (22) 767 7910
Téléphone/Telephone :
Direct : +41 (22) 767 6485
Mobile : +41 (75) 411 0401
E-mail : Alexei.Klimentov@cern.ch

ОТЗЫВ

Члена диссертационного совета на диссертацию Краснова Федора Владимировича на тему: «Методология построения цифрового двойника научно-технического центра в нефтегазовой отрасли», представленную на соискание ученой степени доктора технических наук по специальности 05.13.01 — Системный анализ, управление и обработка информации (технические науки).

Об авторе. Федор Владимирович Краснов является высококвалифицированным специалистом в области системного анализа, управления и обработки информации. Его работы по развитию методов цифровизации в области разведки и добычи углеводородов, разработке вероятностной модели скрытых тем, исследованию командообразования и «созданию» коллектива авторов научных статей получили большую известность в научном сообществе и, что не менее важно, используется на практике Научно Техническими Центрами таких крупных компаний как «Газпромнефть». Докторская диссертация Краснова Ф.В. является не только обобщением ранее выполненных им работ, на практике доказавших их эффективность, но и содержит принципиальные предложения автора по дальнейшему развитию методов и алгоритмов анализа текстовой информации, методики для анализа больших массивов динамических данных, и определения качества автоматизированного перевода научных текстов.

Общая характеристика диссертации.

Цель работы. Автор диссертации определяет три основные цели исследований :

1. Комплексный анализ, диагностика и моделирование социальных процессов в организационной среде;
2. Синтез путей решения проблем обратной связи при прогнозировании развития науки по определенным приоритетным направлениям;

3. Разработка методов сбора и анализа результатов деятельности НТЦ для создания и обучения модели эффективности НТЦ с использованием алгоритмов машинного обучения.

Для достижения сформулированных целей, автор определяет задачи, которые необходимо решить, выделим некоторые из них :

- Определение и разработка методов сбора результатов научной деятельности НТЦ;
- Определение результатов научной деятельности НТЦ, влияющих на научную эффективность НТЦ;
- Разработка интегральных показателей эффективности научной деятельности НТЦ;
- Выбор класса моделей для интегральных показателей эффективности научной деятельности НТЦ;
- Создание модели на основе алгоритмов машинного обучения;
- Апробация модели на действующих НТЦ.

Отметим, что весь спектр задач успешно решен автором в процессе исследования.

Научная новизна диссертации состоит в комплексном подходе к разработке и исследованию алгоритмов сбора цифровых артефактов научной деятельности НТЦ, создании интегральных показателей научной эффективности с целью моделирования на основе алгоритмов машинного обучения.

Структура диссертации и логика изложения. Диссертация состоит из пяти глав (включая Введение и Выводы, главы 1 и 5 соответственно), списка литературы из 319 наименований, списков таблиц и рисунков, листинга программного кода. Полный объем работы составляет 455 страниц.

Глава 1 (Введение) содержит все положения, отвечающие требованиям к общей характеристике диссертации. Этот раздел работы является поучительным, т.к. автор обладает широкой эрудицией в области концепции цифровых двойников и научно-технических центров (НТЦ). Автор детально рассматривает построение цифрового двойника НТЦ как парадигму моделирования и формулирует ключевые вопросы к объекту исследований. После чего автор формулирует прямую и обратную задачу исследований, которые, как правильно постулировано автором, дополняют друг друга. В этой же главе автор рассматривает причины возникновения НТЦ в областях нефти(газо)добычи, переработки и экспорте углеводородов. Впечатляет эрудиция автора при описании истории развития нефтяной и газовой отрасли в России, начиная с конца XVI века до сегодняшнего дня, и роль таких столпов Российской науки, как Д.И.Менделеев, И.М.Губкин и Н.Д.Зелинский. Этот исторический экскурс логически связан с описанием мировых тенденций в области технологий добычи углеводородов. При описании новейших технологий автор справедливо отмечает роль ИТ и цифровизации нефтяной отрасли в целом (от “Цифрового месторождения” до “Цифровой электронной нефтедобывающей компании”). Здесь же автор пишет о проблемах социологического характера («бихейвориально-психологического характера» по определению автора) при внедрении информационных технологий нефтедобывающими компаниями. Автор рассматривает подход к анализу информации в этой отрасли и возможности применения стека технологий и алгоритмов «Больших данных» для обработки и анализа данных терабайтного диапазона. Далее в тексте автор сравнивает отечественные и зарубежные программные пакеты, и дает оценку деятельности НТЦ с точки зрения долгосрочных перспектив. В целом данная глава достаточно богата историческими фактами, ссылками и примерами. Важно отметить описание результатов тематического моделирования, проведенных автором, при анализе статей журнала «Нефтегазовое хозяйство» за период с 2008 по 2016 годы, которые

показали наличие различных тем (в области нефтегазовых технологий) в различных номерах данного издания, и отсутствие «сквозных» тем.

Глава 2 посвящена обзору научной литературы по теме диссертации. Автор рассматривает этапы научного исследования, и соглашается с цитируемыми источниками о том, что написание научной статьи, можно считать конечным результатом научного исследования, а само исследование может быть представлено в виде формализованного процесса, реализуемого участниками научной группы. Изучение процесса коллективного социального взаимодействия является одной из задач автора диссертации. Далее автор рассматривает формальные этапы при публикации научных статей и вопрос о коллективном создании знаний. Риторический вопрос о роли авторов фундаментального курса теоретической физики Л.Д.Ландау и Е.М.Лившица, мне был особенно интересен, как участнику крупных физических экспериментов, где этот вопрос периодически дискутируется. На основе анализа многочисленных источников автор диссертации рассматривает способы анализа человеческого взаимодействия, а также параметры, определяющие продуктивность научного творчества, а также место и роль «слова» (в редакции автора «текста») в научной деятельности. Здесь мысль автора очень близка к мыслям, высказанным выдающимся мыслителем У.Эко в его лекции «от Интернета до Гутенберга» и Ю.М.Лотманом (в статьях по топологии культуры), о том что на основе знаний извлеченных из «текста» принимаются управленческие решения. Как пишет Ф.В.Краснов : «Реальный мир находит свое отражение в текстах при помощи авторов, а процесс анализа текста делает обратное: на основе текстов составляет информацию о реальной природе вещей». Далее автор рассматривает методы обработки текстовой информации, существующие модели и классификацию текстов. Отдельно я бы хотел отметить рассмотрение двуязычного корпуса текстов (когда научная статья изначально опубликована в национальном издании, а потом переведена на английский язык). Это очень интересное исследование, проведенное автором диссертации. Исследовательская гипотеза сформулирована следующим образом : «Для двуязычного корпуса научных статей возможно автоматически выделить пары научных статей, перевод которых сделан с помощью статистического машинного перевода». Результаты исследования приведены автором в последующих главах. Далее при анализе социальных сетей автор делает вывод, что граф соавторства научной статьи представляет собой частный случай социальной сети.

Глава 3 посвящена объекту и методам исследований. В этой главе также определена основная цель диссертации – исследование процесса публикаций. После рассмотрения цифровой экосистемы, автор переходит к рассмотрению системы распространения знаний (СРЗ), в частности, как инструмента, помогающего координировать процессы управления и обмена знаниями в области разведки и добычи нефти внутри группы компаний “Газпромнефть”. Автор описывает объекты, хранящиеся в СРЗ, а также функционал, помогающий проводить сравнительный анализ подходов, технологий и методов, применяемых данной компанией. Рассматривая существующие модели и моделирование социотехнических объектов, автор делает вывод, что в XXI веке произошло качественное развитие компьютерных систем предсказательного моделирования и важной частью предсказательного моделирования является имитационное моделирование. Далее автор переходит к определению модели социальной системы, описанию модели для исследования взаимодействия агентов в процессе создания научных статей (модель реализована автором диссертации) и подтемы (вопросы) для проведения более полного исследования. В этой же главе автор посвящает достаточное время рассмотрению теории имитационного моделирования и принципам построения моделей. Согласно логике изложения, это необходимо для перехода к описанию этапов компьютерного имитационного моделирования и методов сбора данных в ходе имитационного моделирования. В этой же главе, рассмотрена вполне конкретная модель процесса публикации научно-практических статей (определение

термина «научно-практическая» приведено в последующем разделе диссертации, и оно не совсем понятно, видимо, это связано со спецификой работы автора с нефтедобывающими компаниями), также не совсем понятно утверждение Ф.В.Краснова о «размывании» авторского коллектива за счет привлечения в «соавторы людей, для перевода статьи на английский или лоббирования командировки на конференцию» (я не уверен, что знаю о таких примерах в своей научной карьере). Рассматривая основную цель диссертации, автор в доступной форме излагает логический каркас исследования, который включает в себя организационную среду (соавторы и их рукопись), процесс публикации, издателей, показатели продуктивности и результаты публикации. После чего дается определения каждого из объектов, а также алгоритм сбора данных, разработанный автором. Рассматривая «результаты публикации» Ф.В.Краснов предлагает не использовать индекс цитирования, как единственный и/или основной показатель, а добавить «граф соавторства». Этот подход позволяет выделить наиболее значимых ученых, работающих в определенной области знаний (методика графа соавторства описана в следующих разделах). Рассматривая «экосистему научного издательства» автор (называя себя во множественном числе «авторы») описывает разработанный им методический подход для анализа содержимого научного издания и пишет о новизне этого подхода. В последующих разделах автор рассматривает теорию суррогатного моделирования и непараметрические модели в аспекте использования их для исследований в области нефтегазовой отрасли, а также более подробно рассмотривает Байесовские методы для определения параметров НТЦ и наличие Интеллектуального Капитала, как скрытого параметра модели. Важным объектом исследования является моделирование самоорганизующихся команд в научном сообществе. Автор описывает различные мотивы создания таких групп, сравнивая «самоорганизацию групп» с подходом «top bottom» (когда группа создается «сверху», указанием руководства компании или научного эксперимента). Автор описывает «жизненный цикл» таких групп от момента принятия решения о создании неформальной команды, присоединие новых участников к существующей группе, окончательного оформления состава группы, модели компетенций и принятия ключевых решений в группе. Введение понятий полного и остаточного кода, позволили автору построить изменяющийся вектор компетенции команды. Не менее важно моделирование графа соавторства, расмотренная в диссертации, и вопрос о гомогенности комнады исследователей. Раздел 3.11 посвящен анализу текстов, в том числе анализу эмоциональной окраски текстов. Автор успешно разработал методику выбора оптимальной архитектуры и гиперпараметров нейронной сети, «позволяющие обучить классификационную модель на публичном наборе данных, содержащем оценочные суждения, и затем предсказать фрагменты текста из научно-практических статей, содержащие оценочные суждения с заданной степенью точности». Значительная часть исследования связана с методикой сравнения корпусов текстов, в частности изучение когерентности двуязычного корпуса текстов и обнаружения связей между статьями опубликованными на разных языках. Одним из многочисленных достоинств данной работы безусловно является исследование структурных отличий научных статей, возникающие при переводе с русского на английский. Впечатляет количество исследованных документов - 484, по 242 документа на каждом языке.

Глава 4 посвящена апробации результатов исследований. Автором была проанализирована публикационная активность НТЦ «Газпромнефть», проведено измерение ИК НТЦ, моделирование командообразования в научной среде, разработана вероятностная модель скрытых тем на основе анализа публикаций в журнале «Нефтяное хозяйство» в 2008/2016 годах. Представляет особый интерес оценка скрытых направлений научных исследований в нефтегазовой отрасли, проведенная автором. Для оценки использовалось содержание почти 1700 публикаций и методика анализа научных текстов, разработанная автором. Эта же методика была применена для обнаружения негативных отзывов о технологиях, применяемых в нефтегазовой отрасли, в научных публикациях. Дополнительно, экспериментальное

подтверждение методики T4C было успешно продемонстрировано для информации, содержащейся в научных патентах.
В главе 5 приведены основные выводы исследования.

Достоверность и степень обоснованности. Основные результаты работы были представлены на Российских и международных конференциях, и опубликованы в реферируемых научных изданиях.

Текст диссертации хорошо структурирован и содержит ссылки на экспериментальное подтверждение представленных результатов. Сами результаты обоснованы, а исследования выполнены на высоком научном уровне. Некоторые замечания к работе :

1. Несогласование подлежащего и сказуемого, когда подлежащее во множественном числе, а сказуемое в единственном (или наоборот). Данная ошибка встречается довольно часто в первых главах диссертации;
2. Неправильное спряжение глагола;
3. Ошибка в расшифровке названия организации ОПЕК;
4. Отсутствие единообразия при перечислении имен ученых в первой главе диссертации, и опечатки в некоторых из имен;
5. Одновременное использование английской (R&D) и русской (НИР) аббревиатур в тексте диссертации;
6. Англоязычное заимствование (англицизмы), при существовании русского слова (наиболее часто используемое в диссертации слово «топик / topic», в смысле «тема»);
7. Отсутствие единой последовательной нумерации в списке литературы и публикаций автора, что затрудняет поиск библиографических ссылок;
8. Автор не объясняет причины выбора модели Glove, а не, например, BERT при анализе тональности текстов;
9. В некоторых утверждениях автор не приводит ссылки на источник(и), что несколько снижает общее впечатление от безусловно сильной работы (один из примеров в разделе 1.2 «Ряд исследований последнего времени демонстрируют уверенную корреляцию..»)
10. Разделение научных публикаций на : «научно-теоретические», «научно-практические» и «научно-методические», мне кажется не очень удачным, но в данном случае, я полностью доверяю автору, показавшему глубокое знание предметной области в нефтегазовой отрасли.

Все приведенные замечания не являются критичными и не умаляют общего впечатления о диссертационной работе.

Заключение. Значение данной работы состоит в новом подходе на процесс публикации научных статей, что позволило определить показатель продуктивности научного исследования и работы творческого коллектива. Федор Владимирович Краснов разработал алгоритм образования научного соавторства, который может иметь широкое применение. Уже

сегодня алгоритм апробирован на НТЦ «Газпромнефть». Автором предложена и подтверждена гипотеза о возможности машинного выделения эмоционально-окрашенных фрагментов текста научных статей. Это особенно интересно, учитывая «нейтральность» языка академических публикаций. Автор исследовал структурные отличия научных статей, возникающие при переводе с русского языка на английский. Для своих исследований автор использовал методику модального тематического моделирования. Особенностью данного исследования является количество документов - 484 документа, по 242 на каждом языке. Одним из интересных результатов диссертации является точное определение «литературного качества» статьи, в частности, переводная статья, написанная на «псевдоанглийском» языке, может быть точно идентифицирована на основе тематической модели.
Полученные в данной диссертации результаты, безусловно, соответствуют квалификации доктора технических наук по специальности 05.13.01 — Системный анализ, управление и обработка информации (технические науки)

Диссертация Краснова Федора Владимировича на тему: «Методология построения цифрового двойника научно-технического центра в нефтегазовой отрасли» соответствует основным требованиям, установленным Приказом от 01.09.2016 № 6821/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», соискатель Краснов Федор Владимирович заслуживает присуждения ученой степени доктора технических наук по специальности 05.13.01 — Системный анализ, управление и обработка информации (технические науки). Пункт 11 указанного Порядка диссертантом не нарушен.

Член диссертационного совета
Доктор физико-математических наук,

Климентов Алексей Анатольевич

Профессор НИЯУ «МИФИ»
Главный научный сотрудник BNL, Upton, NY, USA
Глава комитета по «компьютингу и модели обработки данных» эксперимента ATLAS CERN

31.10.2020