

ОТЗЫВ

научного руководителя на диссертацию

Шачнева Дмитрия Алексеевича

«Семантические модели классификации и анализа данных в больших информационно-аналитических системах»,

представленную на соискание учёной степени кандидата физико-математических наук по специальности 05.13.17 — теоретические основы информатики

В последние десятилетия активно создаются и развиваются информационно-аналитические системы, в которых хранится, перманентно обновляется и обрабатывается информация о деятельности в различных сферах экономики. Такие системы направлены на автоматизацию процессов управления организациями, на поддержку принятия управленческих решений. Диссертационная работа Шачнева Дмитрия Алексеевича посвящена разработке моделей, алгоритмов и программных средств для интеллектуального анализа данных в таких системах на примере сектора научно-исследовательских, опытно-конструкторских и технологических работ.

Существует большое число востребованных на практике задач, в которых возникает необходимость построения списка объектов заданного типа, относящихся к той или иной предметной области, учёта как степени их соответствия такой области, так и значимости отдельных объектов. Примерами являются задачи поиска отдельных результатов деятельности; поиска экспертов; формирования отчётных материалов; построения рейтингов на уровне работников, коллективов исполнителей или организаций в целом с привязкой к предметной области; создания статистических сведений о деятельности структурных подразделений и организаций. Задача в такой постановке соответствует положениям пп. 1, 2, 5 и 9 областей исследования согласно паспорту специальности 05.13.17 — Теоретические основы информатики.

В первой главе диссертации приведены требования, которые предъявляются к моделям, алгоритмам и их программной реализации, и которыми руководствовался автор в ходе диссертационного исследования. Такие требования затрагивают необходимую функциональность, а также аспекты безопасности и сопровождаемости программного кода.

Во второй главе автор проводит обзор существующих методов для обработки текстовой информации и представления её тематики в формализованном виде. С учётом достоинств и недостатков каждого из известных подходов, для определения степени соответствия объектов предметной области диссертации представлен аппарат тематических портретов, которые можно сопоставить каждому объекту в информационно-аналитической системе для формализованного описания его предметной области. Автором разработаны

функции, позволяющие получить численные характеристики схожести двух тематических портретов. Сформулированы в виде теорем и доказаны свойства этих функций. Проведены эксперименты, подтверждающие их работоспособность и эффективность.

В третьей главе разработан алгоритм для автоматизированного построения таких портретов по косвенным связям объектов в условиях отсутствия непосредственно приписанных объектам предметных областей. Доказательно описаны свойства построенных таким образом тематических портретов.

В четвёртой главе для определения значимости объектов, таких как результаты деятельности, предложена гибкая модель, позволяющая динамически задавать правила для расчёта коэффициентов значимости. Модель позволяет декларативно установить критерии отбора результатов, используемые для вычисления коэффициента значимости показатели объекта и порядок учёта вкладов отдельных авторов. Далее при помощи разработанного автором алгоритма и его программной реализации такие правила преобразуются в запросы к реляционным системам управления базами данных на языке SQL.

В пятой главе, как пример комбинации разработанных методов, автором предложен алгоритм поиска экспертов на основе поискового запроса, состоящего из ключевых слов и рубрик. Сначала выполняется поиск результатов деятельности, далее из множества авторов найденных результатов формируется список возможных экспертов. В научно-технической и технологической области деятельности для расчёта коэффициентов значимости результатов могут быть использованы их библио- и наукометрические показатели. Предложен механизм автоматизированного выявления конфликта интересов. Автором также создана программная реализация модуля поиска экспертов на языке Python и исследована её производительность.

В Заключение сформулированы основные результаты и выводы по итогам исследования, а также отмечены перспективные направления дальнейших исследований.

Представленные в диссертации модели, алгоритмы и их программные реализации прошли апробацию на данных, представленных в информационно-аналитической системе «ИСТИНА», которая в течение многих лет используется в МГУ имени М.В. Ломоносова, а также в 20 других научных центрах РАН и Минздрава РФ. По результатам апробации получен акт о внедрении.

В диссертационной работе использовался аппарат и методы теории графов, математической статистики, комбинаторики.

Достоверность и обоснованность полученных результатов подтверждается:

- наличием математических моделей и математической строгостью доказательств заявленных автором свойств;
- результатами обсуждения основных положений диссертации на научных семинарах, всероссийских и международных конференциях, а также публикациями в рецензируемых научных изданиях, в том числе в журналах, индексируемых в системах Web of Science и RSCI.

Соискатель является сотрудником Лаборатории автоматизации экспериментальных исследований НИИ механики МГУ. В 2015 году закончил аспирантуру механико-математического факультета МГУ. В 2017–18 учебном году получал стипендию Правительства Российской Федерации. Д.А. Шачнев участвует в выполнении научно-исследовательских работ НИИ механики МГУ, в том числе в работах по развитию и сопровождению системы подготовки принятия решений на основе анализа информации о результатах научной, педагогической, инновационной деятельности — ИАС «ИСТИНА». Кроме того, он участвовал в работе научного семинара «Проблемы современных информационно-вычислительных систем».

Все результаты, представленные в диссертации, являются новыми и получены автором самостоятельно. В процессе работы над диссертацией Шачнев Д.А. проявил себя как сложившийся ученый, способный самостоятельно ставить задачи, разрабатывать и обосновывать методы их решения. Полученные результаты находятся на мировом уровне.

С учётом изложенного выше, считаю, что диссертационная работа Д.А. Шачнева соответствует критериям, установленным в «Положении о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова» и рекомендуется к защите в диссертационном совете МГУ.05.01 ФГБОУ ВО МГУ по специальности 05.13.17 — теоретические основы информатики (физико-математические науки).

Научный руководитель, профессор кафедры
вычислительной математики механико-
математического факультета МГУ имени
М.В. Ломоносова, д.ф.-м.н., профессор

В.А. Васенин

Подпись профессора В.А. Васенина удостоверяю

Декан механико-математического
факультета МГУ имени М.В. Ломоносова,
д.ф.-м.н., член-корреспондент РАН,
профессор

А.И. Шафаревич