

ОТЗЫВ
официального оппонента
на диссертацию на соискание ученой степени
кандидата физико-математических наук Шачнева Дмитрия Алексеевича
на тему: «Семантические модели классификации и анализа данных в
больших информационно-аналитических системах»
по специальности 05.13.17 – «Теоретические основы информатики»

Задачи информационного поиска и анализа привлекают внимание как исследователей, так и практиков, занятых разработкой информационных систем на основе вычислительной техники с середины 50-х годов прошлого века. Сложность этих задач связана с невозможностью полной формализации. Наиболее интенсивное развитие этой области происходило в последние три десятилетия в связи с появлением глобальной сети интернет, обеспечившей доступность огромного объема информации, а также в связи с ростом доступной мощности вычислительных систем. Основные трудности при решении подобных задач связаны с выбором метода оценки близости документов и поисковых запросов, при этом критерием качества является соответствие получаемых оценок мнениям экспертов.

В диссертации Дмитрия Алексеевича рассматривается ряд проблем, связанных с задачами информационного поиска и анализа данных, применительно к информационно-аналитическим системам. Особенность таких систем состоит в большом разнообразии типов объектов, зарегистрированных в системе. Основное внимание уделено системам учета и анализа научно-технической деятельности, однако применимость предложенных методов и подходов, по-видимому, может быть значительно более широкой.

Неформально, методы поиска, рассматриваемые в диссертации, состоят в сопоставлении наборов ключевых слов, иерархических классификаций, взаимосвязей между объектами и оценками значимости этих объектов. Для применений в системах учета научной, технической и преподавательской деятельности для оценки значимости предполагается использование

наукометрических показателей объектов. Рассматриваемый подход, однако, не предполагает использование полных текстов документов, написанных на естественном языке, что несколько ограничивает потенциально достижимое качество оценок близости.

Все объекты, информация о которых хранится в системе, описываются введенной в диссертации структурой, которая названа портретом и по существу представляет собой набор ключевых слов, снабженных весами. Применение портретов позволяет применить элементы векторных моделей представления информации, которые, как известно, превосходят булевские модели по качеству получаемых результатов.

Наиболее существенным теоретическим результатом диссертации является определение функции оценки близости, основанной на структурных зависимостях в иерархических классификаторах и косвенной оценки близости значений ключевых слов по близости контекстов, в которых эти ключевые слова встречаются. Таким образом, оценка семантической близости делается на основе применений терминов, а не на основе их семантики. Аналогично, значимость объектов оценивается на основе наукометрических показателей, которые, как известно, также не основаны на семантике и являются косвенными. Анализ отрицательных последствий безудержного и неквалифицированного применения наукометрических показателей для административной оценки исследований далеко выходит как за рамки диссертации, так и за рамки данного отзыва, однако применение подобных методов в рассматриваемой системе может давать и положительные эффекты.

Хорошо известно, что применение иерархических классификаторов может приводить к снижению качества результатов. Для предотвращения подобных эффектов в диссертации вводятся весовые коэффициенты.

Для представления информации в системе применяются концепции семантических сетей, представленных в формате RDF, что дает дополнительные возможности улучшения качества оценок близости за счет учета взаимосвязей между объектами.

Важно отметить также практическую реализацию предложенных в диссертации методов и алгоритмов в системе «Истина», эксплуатируемой в Московском университете.

Как и во всякой большой и значимой работе, в диссертации можно заметить ряд недостатков:

1. Представление материала в некоторых местах вызывает недоумение. Например, одно из важных свойств предлагаемых в диссертации оценок в частном случае сформулировано как «Теорема 1», но обобщения это теоремы на более широкие классы объектов почему-то названы «утверждениями».
2. Понятие обратного предиката, введенное в рамках модели RDF, не дает дополнительной информации и только усложняет модель концептуального уровня. Было бы лучше ввести такие дуги на уровне графа, а не на уровне абстрактной семантической модели.
3. В работе отсутствуют указания на близкое по подходам направление исследований, активно развивающееся в последнее десятилетие: knowledge graph, которые также основаны на подходах RDF к построению семантических сетей.

Указанные замечания не умаляют значимости диссертационного исследования и не влияют на общую положительную оценку работы. Диссертация отвечает требованиям, установленным Московским государственным университетом имени М.В.Ломоносова к работам подобного рода. Содержание диссертации соответствует паспорту специальности 05.13.17 – «Теоретические основы информатики» (по физико-математическим наукам), а также критериям, определенным пп. 2.1-2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В.Ломоносова, а также оформлена, согласно приложениям № 5, 6 Положения о диссертационном совете Московского государственного университета имени М.В.Ломоносова.

Таким образом, соискатель Шачнев Дмитрий Алексеевич заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.17 – «Теоретические основы информатики».

Официальный оппонент:

доктор физико-математических наук, профессор,
профессор департамента информатики
Санкт-Петербургской школы физико-математических наук
Федерального государственного автономного образовательного учреждения
высшего образования «Национальный исследовательский университет
“Высшая школа экономики”»

Новиков Борис Асенович

Контактные данные:

тел.: 7(921)9148534, e-mail: bnovikov@hse.ru

Специальность, по которой официальным оппонентом
защищена диссертация:

05.13.17 – Теоретические основы информатики

Адрес места работы:

194100, Санкт-Петербург, ул. Кантемировская д. 3, корп. 1, лит. А
Национальный исследовательский университет «Высшая школа экономики»,
Санкт-Петербургская школа физико-математических и компьютерных наук
Тел.: +7 (812) 644-59-11 доб. 61578; e-mail: hse@hse.ru