ОТЗЫВ

официального оппонента на диссертационную работу Бондарчука Дмитрия Вадимовича «Алгоритмы интеллектуального поиска на основе метода категориальных векторов», представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 – теоретические основы информатики

Актуальность темы диссертационной работы Д. В. Бондарчука обусловлена постоянным ростом объема текстовой информации, которую необходимо быстро обрабатывать. Требования по времени обработки и объемы информации таковы, что ее сложно обрабатывать ручными способами.

В работе реализованы алгоритмы обработки некого массива текстовых документов. Строится индекс, в котором для каждого документа определяется близость его определенной категории документов. Далее, имея поисковый запрос пользователя, необходимо найти подходящие ему по смыслу документы. Актуально и важно для любого поискового запроса выдавать непустой результат, так как, имея непустой результат, хотя бы отдаленно похожий на искомое, пользователь может понять, как ему актуализировать запрос, а имея пустой результат, он этого сделать не сможет.

Для повышения качества работы автоматических методов имеет также смысл анализировать тексты с учетом структуры естественного языка, полисемии и синонимии, учетом орфографических ошибок и семантических зависимостей между словами. Повышать качество необходимо, так как растет важность тех решений, которые осуществляются по результатам работы алгоритмов.

Целью диссертационного исследования Д. В. Бондарчука является разработка алгоритма интеллектуального анализа текстовых данных, гарантирующего, что пользователь на любой свой запрос получит непустую выборку, отсортированную по степени «полезности».

Для решения этой цели Д. В. Бондарчуком были *поставлены и решены следующие задачи*:

1. Разработан алгоритм интеллектуального анализа текстов, гарантирующего непустой результат независимо от распределения обучающей выборки по категориям. При этом применяется векторная модель представления документа.
2. Разработан метод уточнения весов в векторной модели представления документа с учетом семантической взаимосвязи между термами.
3. Получен ряд теоретических результатов, предназначенных для обоснования работы методов.
4. Проведены эксперименты для: подтверждения работоспособности разработанных методов и алгоритмов; сравнения их результатов, по вре-

мени работы, требуемым ресурсам и качеству, с результатами уже существующих разработок.

Научная новизна исследования: Д.В. Бондарчуком предложены оригинальные методы учета семантической связи между термами, выбора актуальных для решения задачи термов, метод получения гарантированного результата поиска.

Теоретическая ценность заключается в разработанных Д.В. Бондарчуком методах поиска, их теоретическом обосновании с использованием математических моделей и методов.

Достоверность. Предложенные методы и подходы логически строго обоснованы и квалифицированно протестированы на стандартных наборах данных, что обеспечивает достоверность полученных результатов и выводов.

Практическая ценность работы определяется тем, что разработанные методы и алгоритмы могут применяться в интеллектуальных поисковых системах, интернет сервисах, ориентированных на продажу товаров, системах подбора вакансий, системах автоматического формирования персональных рекомендаций.

Результаты прошли апробацию на ряде научных конференций, в том числе, международных. Опубликованы статьи в реферируемых журналах, которые отражают содержание диссертации достаточно полно. Результаты внедрены на нескольких предприятиях.

В качестве **замечаний** имеет смысл отметить:

1. Часто используется термин «семантическое пространство», но его формальное определение отсутствует.
2. Результаты экспериментов даны для небольших документов, вида аннотаций книг, вакансий. Не даны критерии того, каким может быть максимальный размер документа, чтобы его можно было обрабатывать разработанными методами. Неясно, какими будут стабильность, работоспособность алгоритмов и качество результатов, если обработке подвергнутся большие документы (книги, журналы, размером несколько мегабайт).
3. На стр. 85 допущены опечатки при определении формулы similarity (3.1): не хватает скобки в первой строке, напечатано *hyypp* вместо *hyppo*.
4. На стр. 111 формула purity (4.2) определена некорректно с непонятным описанием обозначений.
5. На стр. 115 в таблице даны «Размеры модели представления данных», для разных алгоритмов. Хотелось бы, чтобы структура модели хранения данных была описана, хотя бы для алгоритма, предлагаемого автором.

Указанные замечания не снижают общей значимости исследования и не влияют на общую положительную оценку работы.

Диссертация Д. В. Бондарчука представляет собой законченную научно-квалификационную работу, в которой предложено решение имеющей существенное значение в области информационного поиска задачи

разработки эффективного алгоритма интеллектуального анализа текстовых данных. Текст автореферата соответствует содержанию диссертации.

Считаю, что диссертация Д. В. Бондарчука в полной мере соответствует требованиям, предъявляемым к кандидатским диссертациям, установленным Положением о присуждении ученых степеней, утвержденным постановлением Правительства РФ от 24 сентября 2013 года № 842.

Диссертация соответствует специальности 05.13.17 – теоретические основы информатики.

Дмитрий Вадимович Бондарчук заслуживает присуждения учёной степени кандидата физико-математических наук по специальности 05.13.17 – теоретические основы информатики.

Официальный оппонент
кандидат физико-математических наук
доцент кафедры вычислительной математики и
компьютерных наук,
Федеральное государственное автономное образовательное
учреждение высшего образования «Уральский федеральный
университет имени первого Президента России
Б.Н.Ельцина»
620002, Россия, г. Екатеринбург, ул. Мира, 19.

Веретенников Александр Борисович

22 февраля 2017 г.
E-mail: alexander@veretennikov.ru, AlexanderBorisovich@urfu.ru
Тел.: +7 343 389-94-29.

Подпись Веретенникова А.Б.
удостоверяю
ЗАМЕСТИТЕЛЬ НАЧАЛЬНИКА
УПРАВЛЕНИЯ ПЕРСОНАЛА
22 ФЕВ 2017
С.Ю. Бурдов