

ОТЗЫВ

на автореферат диссертации А.С. Мохова «Метод классификации библиографической информации на основе комбинированных профилей классов с учетом структуры документов», представленной на соискание ученой степени кандидата технических наук по специальности 05.13.01 -системный анализ, управление и обработка информации (в науке и промышленности)

Диссертационная работа А.С. Мохова посвящена разработке математического и программного инструментария с целью повышения эффективности информационного обеспечения научных исследований. Технологии автоматизированного отслеживания и анализа научных публикаций в интересах пользователя дают возможность снизить информационную нагрузку на специалистов и обеспечить их ценными сведениями для проведения научно-исследовательской работы.

Основное внимание в диссертации А.С. Мохова уделяется вопросам повышения точности классификации двуязычных библиографических документов. Для разработки нового метода автором предложено анализировать наличие у библиографических документов определенной (заранее известной) внутренней структуры, которая включает три части: название, аннотацию и ключевые слова. При расчете весов терминов учет их местоположения (наличие-отсутствие в структурных частях документа) позволяет автору повысить точность классификации библиографических текстов. А.С. Моховым предложены новые алгоритмы (UNI5 и UNI6), которые позволяют построить комбинированные профили классов на основе включения в них наиболее информативных русско- и англоязычных терминов. А.С. Мохов разработал методику построения высокоточных классификаторов на основе использования предложенных им алгоритмов и метода, а также формирования на их основе комплекса решающих правил. Сформированный из пяти классификаторов КРП, принимающий решение на основе простого голосования, в ходе экспериментов показал точность более чем на 5 процентов выше по сравнению с наиболее точным индивидуальным классификатором (метод РО-профилей).

По автореферату имеются отдельные замечания.

1. Не совсем понятно, насколько универсальны предлагаемые алгоритмы и возможно ли их применение для проведения классификации, например, новостной информации, имеющей достаточно четкую внутреннюю структуру, а также сообщений электронной почты, не обсуждено, проводились ли исследования на полнотекстовых выборках.

2. Выборки, используемые в исследованиях, описаны довольно кратко. Сложно оценить насколько пересекаются тематики (классы). Хотелось бы видеть примеры составленных выборок.

