

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)

Работа Мохова А.С. посвящена важной теме - классификации текстовых документов. В рамках работы автор делает обзор известных методов, а также на основе комбинирования профилей классов предлагает алгоритмы UNI5 и UNI6, повышающие точность классификации по сравнению с известными методами (например, k-ближайших соседей). Кроме того, автор предлагает метод учета структуры документов - метод Struct. В данном методе предложено придавать различные веса терминам, встречающимся в названиях, аннотациях или ключевых словах. Автором, на основе экспериментов, показано, что терминам из названий целесообразно придавать больший вес, чем ключевым словам и аннотациям, что успешно используется для повышения точности классификации в методе Struct.

Основные научные результаты работы следующие:

1. Предложены алгоритмы UNI, которые, за счет объединения различных подходов выявления информативных терминов, позволяют увеличить точность классификации.
2. Разработан метод Struct, увеличивающий точность классификации за счет учета местонахождения термина в документе. Произведена настройка метода Struct для более точной классификации библиографических описаний, имеющих структуру "название - аннотация - ключевые слова". Оценена вычислительная сложность и даны рекомендации по настройке параметров метода.

Все предложенные в диссертации алгоритмы и методы реализованы в разработанном автором программном комплексе (ПК) "Textcat". ПК был применен в Институте Проблем Химической Физики (ИПХФ) РАН для решения практических задач анализа базы статей, написанных сотрудниками института. По результатам апробации ПК "Textcat" в ИПХФ РАН автором получен акт об использовании разработанных алгоритмов и программ. ПК "Textcat" также был внедрен в учебный процесс кафедры Управления и информатики НИУ "Московский Энергетический Институт", что подтверждается соответствующим актом.

По автореферату имеется ряд **вопросов и замечаний**:

- 1) Почему из известных коэффициентов ассоциативности (КА) выбраны коэффициент Жаккара и коэффициент Сокала-Сниса? Какие другие КА исследовались в диссертации?
- 2) Как и насколько существенно длина профиля класса влияет на точность классификации? Насколько длина профиля изменится в случае проведения классификации одноязычных библиографических описаний?
- 3) Имеются опечатки, в частности в заголовках таблиц на сс. 14 и 17.

В целом, несмотря на сделанные замечания, рассмотренная работа представляет собой законченный научно-квалификационный труд. Автором решена важная для науки и практики задача по увеличению точности классификации двуязычных библиографических документов. Результаты работы опубликованы в 14 статьях, три из которых - в журналах из перечня ВАК, автором также получено свидетельство о регистрации программы для ЭВМ.

Согласно содержанию автореферата, диссертация А.С. Мохова соответствует требованиям ВАК, и соискатель заслуживает присвоения ему ученой степени кандидата технических наук по специальности 05.13.01 – системный анализ, управление и обработка информации (в науке и промышленности).

доцент кафедры
управления и информатики в технических системах
ФГБОУ ВО «МГТУ «СТАНКИН»

к.т.н., доцент

Чеканин Владислав Александрович



Чеканин В.А.

Адрес организации:

127055, Москва, Вавковский пер., 1

тел: (499) 973-30-76

www.stankin.ru

