

## ОТЗЫВ

на диссертацию магистра информационных технологий  
« Анализ пользовательского поведения для уточнения релевантности документа в Web-  
поиске»

Сафоновой Ангилины Владимировны

Тема диссертации весьма интересная – оказывается, что крупные коммерческие поисковики отказываются от попыток победить в конкурентной борьбе, оценивая релевантность документа запросу с учетом контента документа и/или его позиции в Веб-графе (PageRank). Вместо этого проводится глубокий анализ поведения большого числа пользователей, информация о котором хранится в логах, накопленных поисковыми компаниями в течении длительного времени. Таким образом, проблема сводится к разработке метода, позволяющего с высокой вероятностью отгадать, что именно думает пользователь о том или ином документе, базируясь исключительно на его поведении.

Работа, проведенная автором в процессе написания диссертации, просто огромна. Сразу же можно сказать, что и по глубине анализа литературы, и по методике исследования, и по полноте проведенных экспериментов, и по качеству полученных результатов данная диссертация может быть оценена максимально возможной оценкой. Более детальный анализ результатов потребовал бы слишком много места и содержал бы только положительные комментарии за исключением пожалуй части посвященной описанию SVM – приведенное описание совершенно стандартное и от него можно было бы вполне отказаться.

Более полезным и интересным было бы задать автору ряд вопросов и выслушать ответы на защите. Первый вопрос — верно ли утверждение, что в рамках предложенной методики неявно предполагается, что все пользователи могут быть промоделированы одной моделью, т. е. никак не учитывается то, что у разных пользователей могут быть различные идеи о том, что релевантно, а что нерелевантно? Второй вопрос – почему ничего не говорится о семантическом поиске? Быть может это именно тот путь, на котором можно успешно конкурировать, анализируя контент, а не только поведение пользователей? И последний более технический вопрос – верно ли, что при работе с SVM не использовались нелинейные ядра, т. е. заранее предполагалась линейная делимость объектов из разных классов (возможно с относительно небольшим числом ошибок). Если это так, то почему?

В заключение можно еще раз сказать, что диссертация Сафоновой Ангилины Владимировны заслуживает оценки «отлично» и может быть рекомендована к опубликованию, а ее автор заслуживает присвоения степени магистра информационных технологий.

Рецензент,  
к.ф.-м.н., доцент



Добрынин В.Ю.