

## ОТЗЫВ

на автореферат диссертации Смирнова Ивана Валентиновича «Методы разноуровневого анализа текстов на естественном языке и их приложения в системах информационного поиска и психолингвистических исследованиях», представленной на соискание ученой степени доктора технических наук по специальности 1.2.1 «Искусственный интеллект и машинное обучение»

Проблема анализа неструктурированной информации и извлечения знаний из текстов для систем поддержки принятия решений существует давно. На сегодняшний день востребованы инструменты, позволяющие получать такие знания из неструктурированной информации с помощью методов компьютерной лингвистики и интеллектуального анализа текстов. В этих системах представление текста в виде множества слов явно недостаточно, поэтому нужно использовать семантические подходы для структурирования полезной информации, содержащейся в тексте, и разрабатывать новые методы решения этой проблемы.

В работе Смирнова И.В. предложен ряд методов и алгоритмов семантического анализа текстов, предназначенных для извлечения из текста структур, описывающих семантику высказываний. Такие структуры могут использоваться как для представления знаний, выраженных в текстовой форме, так и для вычисления близости двух текстов по смыслу. Предложены также методы дискурсивного анализа, позволяющие работать со структурами, отражающими связность текста. Разработаны методы информационного поиска с применением семантических структур, а также методы извлечения информации из текстов, применимые для автоматизированного формирования баз знаний в виде семантической сети связанных сущностей. Показано также, что семантические структуры позволяют повысить эффективность решения психолингвистических задач. Необходимо отметить, что предложенные в диссертации методы предназначены для обработки текстов на русском языке, для которого аналогичных решений довольно мало.

Практическая значимость работы состоит в применимости разработанных методов к созданию прикладных систем поиска и анализа текстов. Она подтверждается разработкой программных средств текстовой аналитики на базе предложенных в диссертации решений. Указанные программные средства используются как в отраслевых, так и научных организациях, о чем свидетельствуют акты об использовании.

Основные результаты диссертации опубликованы в 51 научной публикации, включая 18 публикаций в изданиях из Перечня ВАК и 21 публикацию в изданиях, индексируемых в БД Scopus и Web of Science. Результаты прошли апробацию на профильных международных конференциях.

В целом работа производит положительное впечатление, однако к автореферату имеются следующие замечания:

1. В автореферате не указано, применим ли предложенный метод извлечения дефиниций из научных текстов к извлечению других сущностей из текстов других жанров.
2. В автореферате отсутствует обоснование выбора реляционно-ситуационной модели для представления семантической структуры текста при том, что известны другие подходы к моделированию семантики текста.

Тем не менее, эти замечания не влияют на общую положительную оценку диссертационной работы, которая выполнена на высоком научном уровне и соответствует критериям, предъявляемым ВАК к докторским диссертациям. Считаю, что Смирнов Иван Валентинович заслуживает присуждения ученой степени доктора технических наук по специальности 1.2.1 – «Искусственный интеллект и машинное обучение».

Татьяна Альбертовна Гаврилова  
доктор технических наук, профессор,  
кафедры информационных технологий в менеджменте  
Высшей школы менеджмента СПбГУ  
Адрес: г. Санкт-Петербург, Волховский пер., д.3.  
тел.: +7 (812) 363-60-00  
E-mail: gavrilova@gsom.spbu.ru

