

О Т З Ы В

на автореферат диссертации Бутенко Юлии Ивановны «Модели и методы автоматической обработки научно-технических текстов в параллельном корпусе», представленную на соискание ученой степени доктора технических наук по специальности 2.3.8 – «информатика и информационные процессы».

Диссертация Ю.И. Бутенко посвящена исследованию и решению комплекса теоретических и прикладных задач, связанных с автоматической обработкой научно-технических текстов при формировании параллельных корпусных ресурсов. Исследования подобного рода проводятся в рамках разработки и создания информационных моделей и методов решения задач обработки научно-технических текстов при создании параллельного корпуса. Подобная проблематика находится на стыке информатики, искусственного интеллекта и компьютерной лингвистики и относится к числу наиболее востребованных направлений развития современных информационных систем.

Содержание автореферата диссертации изложено понятным языком, по прочтении самого автореферата у меня возникло устойчивое впечатление, что он написан специалистом высокого уровня, прекрасно знакомым с текущим положением дел в данной области. Результаты диссертанта, описанные в автореферате впечатляют. Особенно мне хотелось бы отметить следующие.

1. Разработаны теоретические основы обработки научно-технических текстов, которые характерны адаптацией идеи обработки языковых объектов как системы взаимосвязанных компонентов. Результаты подобных разработок вполне могут способствовать развитию эффективных подходов к закреплению смыслового содержания научно-технических текстов.
2. Проведена формализация композиционной структуры документов различного типа – научных статей, учебно-научных изданий и нормативной документации – с последующим использованием полученных моделей в задачах автоматической разметки и выравнивания параллельных текстов.
3. Выделены структурные модели русско- и англоязычных многокомпонентных терминов, а также предложен метод разметки этих терминов в корпусе научно-технических текстов.
4. Предложен комплексный подход, основанный на разноуровневом анализе научно-технических текстов.
5. Предложены методы идентификации машинно-сгенерированных и машинно-переведенных текстов.
6. Разработан прообраз системы управления корпусными данными, новизна которого состоит в возможности управлять подобными данными на разных этапах их обработки, а также формировать различные наборы данных для машинного обучения.

Эти и другие результаты диссертации, нашедшие отражение в автореферате, заслуживают самой высокой оценки. Результаты диссертации опубликованы в отечественных и зарубежных журналах высокого рейтинга. Публикации по теме диссертации содержат описание примененной методики исследования и отражают основные полученные результаты, изложенные в автореферате, а Юлия Ивановна Бутенко, на мой взгляд, безусловно заслуживает присвоения ученой степени доктора технических наук по специальности «2.3.8 – Информация и информационные процессы».

6 февраля 2026 г.

Ведущий научный сотрудник

Математического института имени В.А. Стеклова РАН

(119991, Москва, ул. Губкина, д. 8, тел. +7(495) 984 81 41,

web-сайт www.mi.ras.ru)

доктор физико-математических наук, профессор

(тел. +7 (495) 984 81 41 * 37 36, e-mail: ilichev@mi-ras.ru)



А. Т. Ильичев

Подпись А. Т. Ильичева заверяю:

Ученый секретарь Математического ин-та им. В.А.Стеклова,

к.ф.-м.н



С.А. Поликарпов