

**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФИЛОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ**

Направление «Лингвистика»

Образовательная программа «Прикладная и экспериментальная лингвистика»
Профиль «Компьютерная лингвистика и интеллектуальные технологии»

РЕЦЕНЗИЯ

на выпускную квалификационную работу (магистерскую диссертацию) студента
Дубовик Анны Романовны

на тему: «Автоматическое определение стилистической принадлежности текстов по их
статистическим параметрам»

Актуальность ВКР, новизна

- В связи с постоянным ростом объемов текстовой информации и необходимостью ее автоматической обработки и классификации работа, безусловно, актуальна. Работу отличает определенная новизна, в частности, в области автоматического выделения функциональных стилей в русском языке, поскольку алгоритм, использовавшийся в работе, является авторским.

Оценка содержания работы

- Содержание работы соответствует уровню магистерской диссертации. Использованная литература на двух языках в полной мере отвечает требованиям теоретической базы для данного исследования. Все задачи исследования успешно решены, цель исследования (разработка инструмента автоматического определения стилистической принадлежности текстов) достигнута.

Положительные стороны ВКР

- В теоретической части подробно описаны различные критерии выделения функциональных стилей в русском языке, а также основные проблемы, связанные с этим;
- В главах 2 и 3 приводится подробное описание используемого программного обеспечения;
- В главе 3 производится скрупулёзный анализ полученных статистических данных.

Практическое значение работы

- В ходе исследования были выделены отличительные характеристики для каждого функционального стиля, значимость которых для выделения функциональных стилей была подтверждена статистическими данными;
- Созданная в рамках работы программа может быть использована для дальнейших исследований в области стилистики и определения функциональных стилей текстов.

Недостатки и замечания по ВКР

- Характеристики научного и художественного стилей описаны слишком кратко (по сравнению с публицистическим и деловым стилями); так, на стр. 14 в списке

характеристик художественных текстов упоминается «употребление преимущественно семантически конкретных существительных» без каких-либо дополнительных пояснений. Хотелось бы, чтобы характеристики были проиллюстрированы примерами (как, например, на стр. 16 и 17);

- Не совсем точно оформлены ссылки в контексте списков (стр. 6, 7, 21, 23): если в списке приводится информация из источника, то ссылка должна находиться в конце списка.

Вопросы дискуссионного характера

- В качестве материала для исследования автор выбрал четыре монотематических корпуса, представляющих различные функциональные стили. Почему автор выбрал столь близкие по тематике тексты (ракетостроение, радиоэлектроника и техника)? Можно предположить, что при работе с неоднородным корпусом точность алгоритма фильтрации текстов целевых групп должна быть выше той, которая установлена в результате экспериментов, что свидетельствует о надежности разработанного алгоритма.
- Процедура автоматической диагностики стиля, разработанная автором, опирается на набор информативных параметров, среди которых есть морфосинтаксические параметры (например, доля слов различных частей речи, доля конструкций различных типов и т.п.) и графематические параметры (средняя длина слов, предложений). Хотелось бы узнать у автора, насколько равноценны эти параметры, можно ли их ранжировать по важности (или весу)?
- В результате автоматической обработки четырех монотематических корпусов, отличающихся стилистической принадлежностью, автор получил данные о частотности слов разных частей речи. Интересно узнать, насколько эти данные коррелируют с частеречной статистикой в доступных корпусах русского языка, в частности, в НКРЯ (поскольку есть соответствующие словари).

Данные соображения носят исключительно дискуссионный характер и призваны подчеркнуть достоинства проведенного исследования.

Таким образом, работа заслуживает оценки «отлично», а студент Анна Романовна Дубовик – присвоения квалификации магистра.

Рецензент

Александра Владимировна Коновалова,
ООО «ПРОМТ», младший лингвист

