

М. А. Зильберглей, д-р хим. наук, доц.;  
Ю. Ф. Шпаковский, доц., канд. филол. наук;  
М. М. Невдах, канд. техн. наук  
(ИОНХ НАН Беларуси, БГТУ, ИООО «ЭПАМ Систем»,  
г. Минск)

## ОБЪЕКТИВНАЯ ОЦЕНКА КАЧЕСТВА АВТОРСКИХ МАТЕРИАЛОВ

Современные информационные технологии открывают новые пути развития для статистической обработки текстов. В настоящее время появилась возможность обрабатывать большие объемы текстовой информации за приемлемое время. В данном случае под обработкой понимаются такие действия как разбиение исходного текста на структурные единицы, выделение морфологических и синтаксических признаков единиц текста, количественная обработка результатов.

На кафедре редакционно-издательских технологий Белорусского государственного технологического университета активно ведется разработка автоматизированных систем, помогающих выполнять информационные, логические, аналитические и другие задачи. С этой целью решаются следующие задачи:

- статистическая обработка и анализ различных видов текстов;
- разработка комплекса компьютерных программ для автоматизации анализа (морфологического, синтаксического и др.) текстов;
- анализ существующих методов для определения качества текстов для читателей, проведение экспериментов, обработка и анализ результатов;
- оценка качества текстов для будущих читателей;
- разработка программных средств для распознавания материалов, оптимальных для восприятия читателями;
- разработка универсальной программной среды, предоставляющей возможность обрабатывать и анализировать текст, оценивать его качество для читателя.

Одно из первых отечественных исследований в области оценки качества авторских материалов для будущих читателей проведено в начале 2000-х гг. [1]. Автором работы была предложена концепция определения реальной трудности текста (его качества) для конкретной категории читателей. Данная концепция была взята за основу в более поздних исследованиях и включала следующие этапы:

- поиск объективных критериев, определяющих качество текста отмеченной категорией читателей;
- выявление текстовых параметров, величины которых позволяют оценить качество текста;
- разработка надежных способов измерения величин текстовых параметров;
- определение однозначной функциональной зависимости между величинами отмеченных текстовых параметров и качеством текста.

В работе [1] на первом этапе были выделены морфологические и синтаксические параметры (длина слов и предложений, разнообразие словаря, трудность межсловных связей, процент цитат, иностранных слов и др.) и разработаны способы измерения их величин. Всего было выделено и рассчитано 83 признака.

На втором этапе были найдены объективные критерии, определяющие качество текста. Среди них: процент неправильно заполненных пропусков и время работы с текстом (с использованием методики дополнения), средняя оценка трудности текста и время работы с ним (с использованием экспертных оценок).

На третьем этапе был установлен характер взаимосвязей и выявлена степень воздействия различных факторов на качество материала (материалом для экспериментов послужили тексты из учебных изданий по химии для высших учебных заведений). На основе выявленных в экспериментах факторов трудности текста стало возможным выразить связь между результативным и факторными признаками с помощью специального уравнения линейной множественной регрессии (формулы читабельности), которое позволяет объективно

оценивать качество авторских материалов в автоматизированном режиме.

В работе [2] под руководством проф. М. А. Зильберглайта была разработана модель контроля качества изданий на допечатной стадии полиграфического производства. Для создания модели были решены следующие задачи:

– проведены классификация, анализ и отбор методов проверки качества печатного материала на допечатной стадии полиграфического производства (экспериментальными материалами послужили издательские оригиналы для вузов по философии и экономической теории), установлены объективные показатели качества исследуемых объектов выборки (использовались методика дополнения, метод балльных оценок и метод парных сравнений);

– определены и измерены количественные характеристики исходного материала и экспериментальным путем выявлена их связь с качеством; снижена размерность признакового пространства методами многомерного статистического анализа (кластерный и факторный анализ, метод корреляционных плеяд и вроцлавской таксономии, многомерное шкалирование) и выявлены наиболее информативные показатели (связаны с длиной слов и предложений, числом предикативных ядер, длиной фраз, разнообразием словаря, абстрактностью изложения и др.);

– на последнем этапе на основе экспертных данных и информативных признаков с помощью дискриминантного анализа разработана модель процесса в виде набора классифицирующих функций для автоматизированного контроля качества изданий и разработка специализированного программного обеспечения для внедрения в существующий технологический процесс издательских предприятий.

В соответствии с требованиями к программным средствам был разработан программный продукт «Readability analysis», предназначенный для автоматизированного контроля качества материалов на подготовительной стадии.

Практическая значимость полученных на кафедре редакционно-издательских технологий результатов связана с тем, что они могут быть использованы в редакционно-издательской деятельности при подготовке изданий к печати. Объективный анализ качества текста на стадии его подготовки и дальнейшее усовершенствование материала позволят привести уровень сложности текста в соответствие со способностями читательской группы.

Результаты проведенных исследований дают возможность продолжить автоматизацию редакционно-издательского процесса. Полная или частичная замена человека специализированной системой позволит добиться не только невозможного для человека быстродействия, но и необходимого качества изданий благодаря объективной оценке качества текста на основе его информационных характеристик, полученных в опоре на восприятие читателей.

Результаты работы внедрены в практику ряда отечественных издательств Республики Беларусь, а также в учебный процесс Белорусского государственного технологического университета.

Работа по разработке автоматизированных систем будет продолжена на кафедре. Наиболее актуальными задачами являются: поиск новых параметров, наиболее статистически значимых для решения поставленной цели; поиск эффективных методов для анализа статистических параметров; выявление наиболее действенных методов определения качества текстов; поиск методов для анализа взаимосвязи между качеством текста и его статистическими параметрами.

## ЛИТЕРАТУРА

1. Шпаковский, Ю. Ф. Оценка трудности восприятия и оптимизация сложности учебного текста (на материале текстов по химии): дисс... канд. филол. наук / Ю. Ф. Шпаковский. — Минск, 2007.
2. Невдах, М. М. Модель контроля качества подготовительной стадии издательско-полиграфического производства: дисс... канд. техн. наук / М. М. Невдах. — Минск, 2014.