

ИСПОЛЬЗОВАНИЕ МОРФОЛОГИЧЕСКИХ ХАРАКТЕРИСТИК СЛОВ ТЕКСТА В КАЧЕСТВЕ КЛАССИФИКАЦИОННЫХ ПРИЗНАКОВ

Аннотация. В статье рассматриваются базовые единицы автоматической обработки текстовой информации на примере автоматической классификации русскоязычной художественной литературы по ее возрастной аудитории. На основе экспериментальных данных мы заключили, что морфологические характеристики лексем текста сами по себе не являются подходящим основанием для классификации.

Ключевые слова: автоматическая обработка текстов; классификация текстов; русский язык.

Введение

Работа посвящена выявлению базовых признаков для автоматической обработки неструктурированной текстовой информации на русском языке на примере автоматической классификации художественной литературы по ее возрастной аудитории. Эта задача актуальна в первую очередь для решения проблемы оптимизации информационного поиска в сети Интернет и хранилищах электронных документов. Визуально определить возраст потенциального адресата текста обычно несложно, трудность заключается в поиске автоматического алгоритма решения задачи.

1. Постановка задачи

В ряде работ [1,2] в качестве отличительной характеристики текстов, предназначенных для младшей возрастной аудитории, указывается ограниченное использование автором функционально-смыслового типа речи описание. Под типом речи [3] понимается способ изложения, избираемый автором и ориентированный на одну из задач: статически изобразить действительность, описать ее; динамически отразить действительность, рассказать о ней; отра-

зить причинно-следственные связи явлений действительности. В соответствии с этими целями коммуникации выделяются три основных типа речи: описание, повествование, рассуждение.

В данной статье мы оценим возможность использования морфологических характеристик описания в качестве основания для классификации текстов по их возрастной аудитории.

2. Корпус

Для проведения эксперимента мы использовали базу Национального корпуса русского языка [4]. Корпус состоит из заведомо качественных и максимально разнообразных текстов на русском языке с известным жанром. Мы работали с двумя выборками — художественными текстами различных жанров (историческая проза, приключения, документальная проза и т. д., кроме детской литературы, — всего 5 902 документа, 9 332 659 предложений, 94 538 056 слов) и детской литературой (всего 632 документа, 547 735 предложений, 4 742 627 слов).

3. Признаки и эксперимент

Функционально-смысловой тип речи описание [3] отличается от других типов речи тем, что оно дает представление о каком-либо предмете или явлении перечислением их признаков и свойств. Ведущую роль в описании играют прилагательные, обеспечивающие выразительность и наглядность изображения. Описание как способ изложения противоположно повествованию и направлено на статическое отображение действительности, что выражается в меньшей в сравнении с другими типами речи частоте использования глаголов и глагольных форм.

В связи с этим в выборках проводился поиск следующих характеристик и их последующее сравнение:

- число глаголов;
- число особых глагольных форм (деепричастия, причастия);
- число прилагательных.

В соответствии с развернутым определением [3] было выдвинуто предположение о том, что в текстах выборки 1 (художественная литература, кроме детской) доля глаголов и особых глагольных форм должна быть выше, что в текстах выборки 2 (детская литера-

тура). И напротив — доля прилагательных в текстах выборки 2 будет превышать соответствующий показатель в выборке 1.

4. Результаты и обсуждение

Значения итоговых показателей представлены в таблице 1. Для каждой выборки и характеристики представлены количества вхождений частей речи и их частота:

$$F_i = \frac{N_i}{N}, \quad (1)$$

где N — общее число слов в выборке, N_i — число вхождений части речи в соответствии с результатами поиска.

Таблица 1

Показатели, полученные на основе выборок

Выборка	Число слов (N)	Число вхождений					
		Глаголы		Особые формы глагола		Прилагательные	
		N_1	F_1	N_2	F_2	N_3	F_3
Выборка 1	94 538 056	18 956 913	0,2	2 858 277	0,03	10 209 748	0,11
Выборка 2	4 742 627	999 715	0,21	118961	0,025	492 593	0,1

Полученные значения позволяют увидеть, что предположение, выдвинутое в пункте 3, является верным. В соответствии с итоговыми показателями частота глаголов в выборке 1 составляет 0,2 (то есть 20%), особых глагольных форм — 0,03, прилагательных — 0,11. При этом частоты для выборки 2 — 0,21, 0,025 и 0,1 соответственно.

Несмотря на полученное подтверждение, различие в полученных значениях невелико, что не позволяет использовать данные характеристики в качестве классификационного признака.

Заключение

Хотя полученные нами показатели соответствуют теоретическим утверждениям об использовании частей речи в различных функционально-смысловых типах речи, разница в этих показателях

при сравнении выборок является незначительной (в нашем случае — меньше 1%). Подводя итог, можно сделать вывод о том, что морфологические характеристики лексем текста сами по себе не являются подходящим основанием для классификации.

СПИСОК ЛИТЕРАТУРЫ

1. Алексеева М. И. Редактирование детской литературы. Московский государственный университет. (2009)
2. Здир В. В. Возраст читателя и специфика произведений для детей // Практическая психология и логопедия. 2009. № 4. С. 44-47.
3. Культура письменной речи — gramma.ru.
4. Национальный корпус русского языка — ruscorpora.ru.

И. И. ГРИГОРЬЕВА, А. С. ОСКОЛКОВ

Тюменский государственный университет

УДК657.1.011.56

РАЗРАБОТКА ПОДСИСТЕМЫ РАСЧЕТА РЕЗЕРВА ОТПУСКОВ В «1С: ПРЕДПРИЯТИЕ 8»

Аннотация. В статье описывается доработка типовой конфигурации «Зарплата и управление персоналом» на базе «1С:Предприятия 8» для расчета резерва отпусков.

Ключевые слова. Резерв отпусков, 1С: Предприятие 8, Зарплата и управление персоналом, Бухгалтерия.

В Российской Федерации на всех предприятиях с целью наблюдения, измерения, регистрации, обработки и передачи информации в стоимостной оценке об имуществе, источниках его формирования, обязательствах и хозяйственных операциях ведется учет хозяйственной деятельности, называемый бухгалтерским учетом [1]. Одной из основных задач бухгалтерского учета является выявление внутрихозяйственных резервов обеспечения финансовой устойчивости организации [2].

Резерв отпусков — это резервный фонд, создаваемый организацией для покрытия своих убытков связанных с выплатой отпусков