



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н. Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н. Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К КУРСОВОЙ РАБОТЕ

НА ТЕМУ:

Разработка базы данных для АРМ разметчика параллельного
корпуса технических текстов.

Студент ИУ7-64Б
(Группа)

(Подпись, дата) К. А. Рунов
(И. О. Фамилия)

Руководитель курсовой работы

(Подпись, дата) Ю. В. Строганов
(И. О. Фамилия)

2024 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
1 Аналитическая часть	5
1.1 Корпуса текстов	5
1.1.1 Применения	5
1.1.2 Типы и классификация	5
1.1.3 Устройство	6
1.2 Тексты	6
1.2.1 Виды текстовых разметок	6
1.2.2 Структура технических текстов и проблемы	6
1.2.3 Проблема терминов	6
1.3 Разметки текстов	7
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	8

ВВЕДЕНИЕ

1 Аналитическая часть

В данной части вводятся понятия корпусов текстов и текстовых разметок. Анализируются основные виды текстовых корпусов, текстовых разметок и их структура. Описывается структура технических текстов, и рассматривается связанная с ними проблематика.

1.1 Корпуса текстов

По Т. МакЭнери и Э. Вилсону, под корпусом текстов понимается собрание языковых фрагментов, отобранных в соответствии с четкими языковыми критериями для использования в качестве модели языка [1].

1.1.1 Применения

Корпуса текстов используются для различных лингвистических, образовательных и технологических целей, например, для выявления закономерностей в языке, анализа структуры технических текстов, обработки естественного языка, составления словарей или интерактивного обучения языку.

1.1.2 Типы и классификация

Различают несколько типов текстовых корпусов по степени «универсальности» [1]:

- 1) корпуса первого типа — универсальные, отражают в себе все многообразие речевой деятельности;
- 2) корпуса второго типа — специфичные, создаваемые для специальной цели, отражают некоторое языковое или культурное явление в общественной речевой практике, например, корпус пословиц или анекдотов;
- 3) корпуса третьего типа — специфичные, создаваемые для решения специальной задачи, например, для отладки систем машинного перевода.

Текстовые корпуса различают и по другим признакам. Некоторые из них представлены ниже, в таблице 1.

Таблица 1 – Классификация корпусов [1, с. 16]

Признак	Тип корпуса
Цель	многоцелевые, специализированные
Тип языковых данных	письменные, устные (речевые), смешанные
«Литературность»	литературные, диалектные, разговорные, терминологические, смешанные
Жанр	литературные, фольклорные, драматургические, публицистические
Назначение	исследовательские, иллюстративные
Динамичность	динамические, статические
Разметка	размеченные, неразмеченные
Характер разметки	морфологические, синтаксические, семантические, анафорические и т. д.
Доступность	свободно доступные, коммерческие, закрытые
Объем текстов	полнотекстовые, «фрагментнотекстовые»

1.1.3 Устройство

1.2 Тексты

1.2.1 Виды текстовых разметок

1.2.2 Структура технических текстов и проблемы

1.2.3 Проблема терминов

Терминология русского языка складывалась и продолжает складываться на основе англоязычной терминологии. Иногда получается, что давно существующий в русском языке термин приобретает новое значение, заимствованное из английского. Так, например, слово «корпус» приобрело новое значение от английского «corpus» — сборник текстов [1]. В связи с этим, при разметке текстов, возникают трудности при сопоставлении терминов из одного языка с терминами другого.

Также, в связи с языковыми и культурными различиями, случается, что какой-то термин из одного языка не имеет точного эквивалента в дру-

гом. Например, не имеет точного эквивалента в русском языке японский термин 侘寂 (Wabi-Sabi), который означает умение восхищаться чем-то несовершенным. Или наоборот, русский «кипяток» нельзя перевести на английский одним словом.

1.3 Разметки текстов

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика: Учебник для студентов направления «Лингвистика». 2-е изд., перераб. и дополн., – СПб.: СПбГУ. РИО. Филологический факультет, 2013. – 148 с.