Отчет 11 задание

Регулярные выражения:

i)С латинскими буквами и цифрами: \b[a-zA-Z0-9]+\b

j) Слова с хотя бы одной латинской буквой: \b\w\*[a-zA-Z]\w\*\b

l) Листовые теги (только текст внутри): <(\w+)>([^<]+)</\1>

Ответы на вопросы (m):

1. **@** в строке — позволяет писать литералы без экранирования
2. **\w\*** — 0+ словарных символов, \W\* — 0+ несловарных.
3. **[^<>]\***— любые символы, кроме **<** и **>**
4. Именованная группа: **(?<name>...)**
5. Использованные элементы:
6. Исчислители: **+, \***
7. Классы символов: **\w, \s, \d**
8. Обычные символы: буквально **., ,, -**
9. Обратные ссылки: **\1**
10. Подстановки: **$1, $2**

Решние практической задачи состояло из:

Дефект 1: Пустые теги форматирования  
Некоторые теги форматирования (например, <bold>, <italic>) содержали только пробельные символы или знаки препинания, что делало их бесполезными и могло вызывать ошибки при дальнейшей обработкею.

Использовано следующее регулярное выражение:

<(\w+)>\s\*[.,;:?!\-\—]\*\s\*</\1>

Дефект 2: Лишние пробелы перед знаками препинания

В тексте встречались лишние пробелы перед знаками препинания (например, слово , слово.), что нарушает правила оформления текста.

Использовано регулярное выражение:

\s+([.,;:?!\-\—])

4. Сомнительные случаи

Несмотря на то, что большинство найденных фрагментов были явными дефектами, были выявлены сомнительные случаи , которые могут быть как ошибками, так и допустимыми исключениями.

Примеры сомнительных фрагментов:

<bold> —</bold> Такой тег содержит длинное тире и пробел перед ним

* Возможно, это опечатка или неверная расстановка знакоd

<italic> ;</italic>

* Содержит точку с запятой после пробела. Не несёт смысловой нагрузки, скорее всего — ошибка форматирования.

<main-word><bold>АБВГДЕЖЗ</bold></main-word>

<article><bold> </bold>Какое-то описание слова.</article>

* Тег <bold> внутри <article> пустой — очевидная ошибка форматирования.