

Title Utils Rule

DocScanner AI Writing Rule Documentation

Rule Overview

Name: Title Utils

File: title_utils.py

Description: Implements Title Utils checking and validation

Purpose: Ensures compliance with Title Utils guidelines

Functions

is_title_or_heading(text, original_html)

Determine if a given text is likely a title or heading. Args: text (str): The text to check original_html (str): Original HTML content for context Returns: bool: True if text appears to be a title/heading

extract_titles_from_html(html_content)

Extract all titles/headings from HTML content. Args: html_content (str): HTML content Returns: list: List of title texts

Regular Expression Patterns

```
) headings = soup.find_all([
^#\s+.\+$
^(chapter|section|part|appendix|figure|table)\s+\d+
^\d+\.\(\d+\.\)*\s+[A-Z]
^(introduction|overview|conclusion|summary|getting
started|installation|configuration)
```

Code Examples

```
Title Detection Utilities Helper functions to identify titles and headings in
content
import re from bs4 import BeautifulSoup def is_title_or_heading(text,
original_html=None):
s in HTML heading tags if original_html: soup = BeautifulSoup(original_html,
```

Dependencies

- bs4.BeautifulSoup
- re

Implementation Notes

- Check if it's in HTML heading tags
- Also check if the text is part of a heading (in case of processing differences)
- Check for markdown heading patterns
- +\s+.\+\$', text):
- Check for typical title characteristics

Generated on 2025-10-17 18:29:07
DocScanner AI - Writing Quality Enhancement System