

Long Sentence Rule

DocScanner AI Writing Rule Documentation

Rule Overview

Name: Long Sentence

File: long_sentence.py

Description: Identifies sentences that exceed recommended length limits

Purpose: Enhance readability by encouraging concise sentences

Functions

__get_nlp()

check(content)

Checks Performed

- Long Sentence Detection

Regular Expression Patterns

```
) text_content = soup.get_text() try: nlp_model = _get_nlp() doc =
nlp_model(text_content) except Exception as e: return [] # Return empty if spaCy
fails # Flag long sentences (>25 words) - exclude titles and markdown tables for
sent in doc.sents: # Skip if this appears to be a title or heading if
TITLE_UTILS_AVAILABLE and is_title_or_heading(sent.text.strip(), content): continue
# Skip markdown table separator rows (| --- | --- | etc.) sent_text =
sent.text.strip() if re.match(r
^\\.|.*\\.|.*\\|
```

Code Examples

```
RAG helper not available for {__name__} - using basic rules
in sent_text or re.match(r
Consider breaking this long sentence into shorter ones:
```

Dependencies

- logging
- title_utils.is_title_or_heading
- re
- spacy
- rag_rule_helper.check_with_rag
- html

- bs4.BeautifulSoup

Implementation Notes

- Load spaCy model lazily to avoid startup issues
- Increase max_length to handle large documents
- Return empty if spaCy fails
- Flag long sentences (>25 words) - exclude titles and markdown tables
- Skip if this appears to be a title or heading

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