Providing content to AI engines using DITA OT plugins
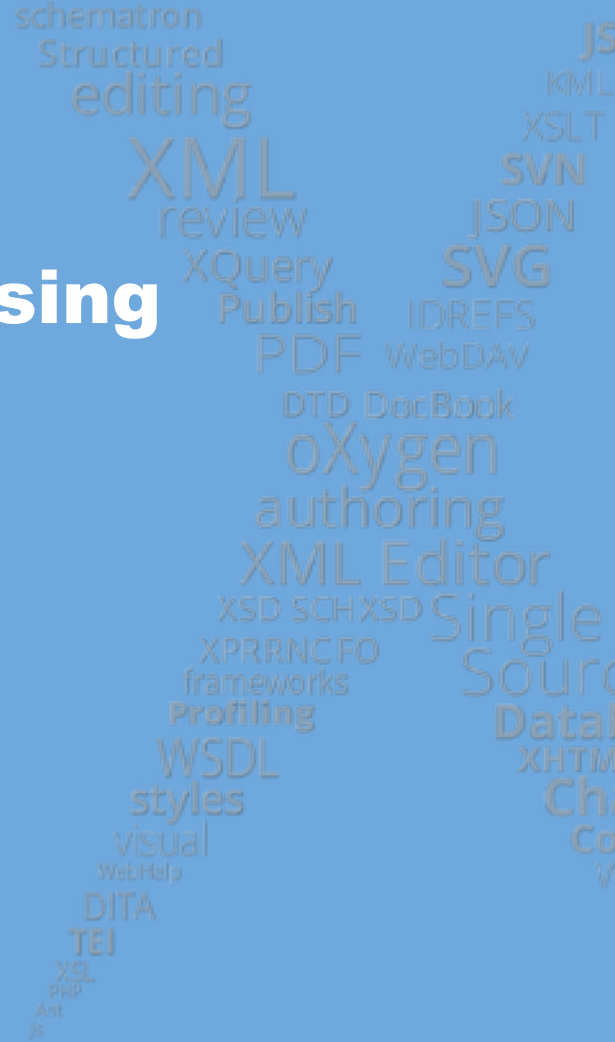
Radu Coravu, Syncro Soft

radu_coravu@oxygenxml.com

Bio

- Software engineer working for Oxygen XML.

- DITA XML expert.

- Contributor in the DITA OT project.

- Lots of DITA XML related technical support.

- Small articles, blog posts, open source projects, documentation updates.

- Working on AI-based DITA XML editing tools.

Radu Coravu - Bio

Overview

Proof of concept implementation of various AI-related features as DITA Open Toolkit plugins

(1) Apply AI Checks on DITA XML Conter

- DITA OT Plugin which pre-processes the DITA XML content (resolves content and key references, applies filters).

- A prompt is applied on each pre-processed DITA XML file and a final report is generated.
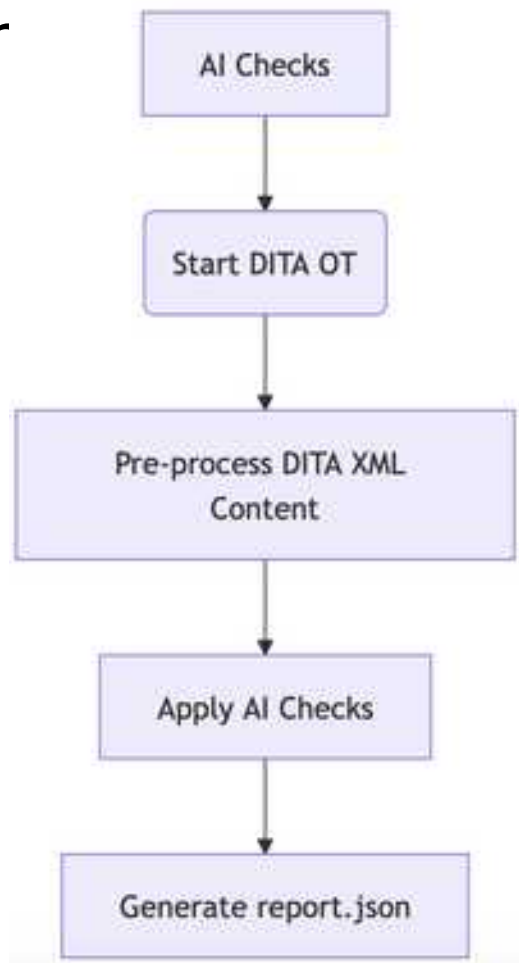
AI Checks Example – Grammar and Spell Check

file:/Users/raducoravu/eclipse–workspace/eXml/samples/dita/flowers/topics/flowers/lilac.dita:
    errors:
        line: 29
        column: 94
        error: use of 'fallacy' in context may require a clearer explanation or could be a misuse of the term given
        the context.
    errors:
        line: 43
        column: 53
        error: Typographical error: 'encurved' should be 'curved' or revised for clarity, as 'encurved' is not a
        standard English term.
file:/Users/raducoravu/eclipse–workspace/eXml/samples/dita/flowers/topics/care.dita:
    errors:
        type: spelling
        message: Mispelled word: Preparaton
        suggestions:

(2) Produce Fine-tune Dataset from DITA Style Guide

- DITA OT Plugin which publishes DITA to Markdown and then creates of question/answer pairs with AI generated examples based on the Markdown content.

- The question/answer pairs are grouped inside a JSONL file which can be used to fine tune an OpenAI model.
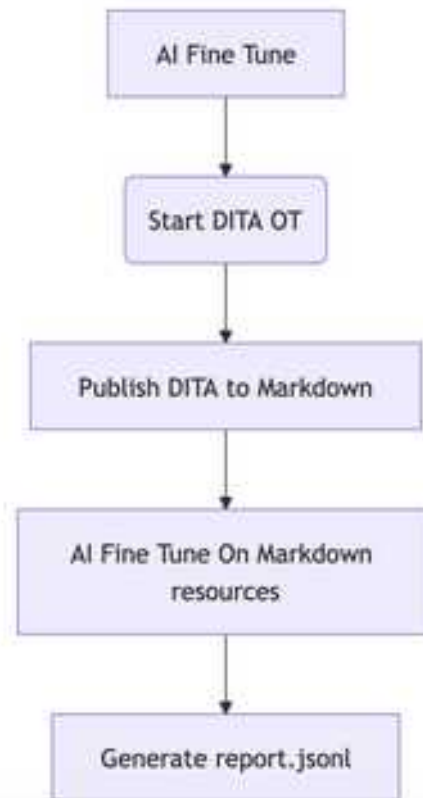
Fine-tune

Use of Specific Words (Style Rules)

The following ▷*Style Rules*◁ provide guidelines for the use of particular words:

▷▷*"Authors"*◁◁ vs. ▷▷*"Writers"*◁◁
    ▷▷**Rule:**◁◁ Simply as a matter of company preference, use ▷*"content authors"*◁ instead of ▷*"technical writers"*◁.
▷▷*"Combo Box"*◁◁ vs. ▷▷*"Combobox"*◁◁
    ▷▷**Rule:**◁◁ Use ▷*"combo box"*◁ instead of ▷*"combobox"*◁.

```
{
  messages:
    role: system
    content: You are a technical writing expert who applies the style guide rules.
  messages:
    role: user
    content: Should I use 'content authors' or 'technical writers'?
  messages:
    role: assistant
    content: You should use 'content authors'.
    Examples:
    Invalid usage: Technical writers create the documentation.
    Valid usage: Content authors create the documentation.
}
```

Fine-tune

Fine-tune

(3) Upload Data to OpenAI Assistant Vector Store

- DITA OT Plugin which publishes DITA to Markdown a
  then uploads the content to an AI Assistant Vector
  Store.
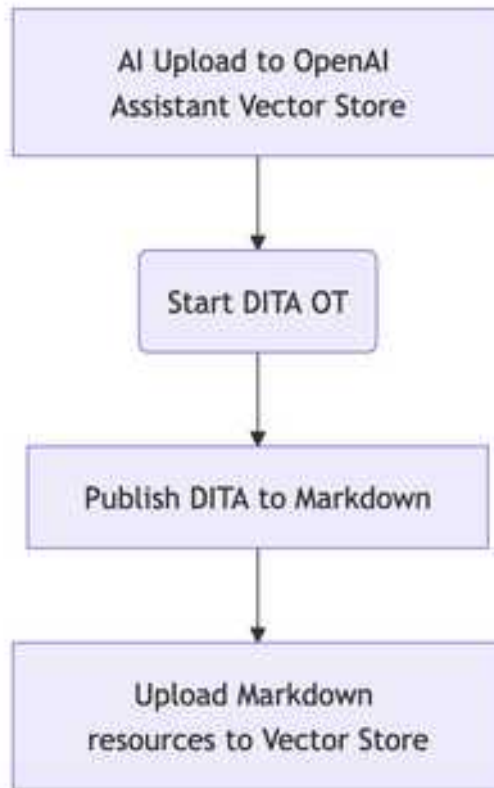
Vector Store Example – Assistant based on User's Guide

(4) AI Preprocess

- DITA OT Plugin which applies modifications using the pre-processing stage.
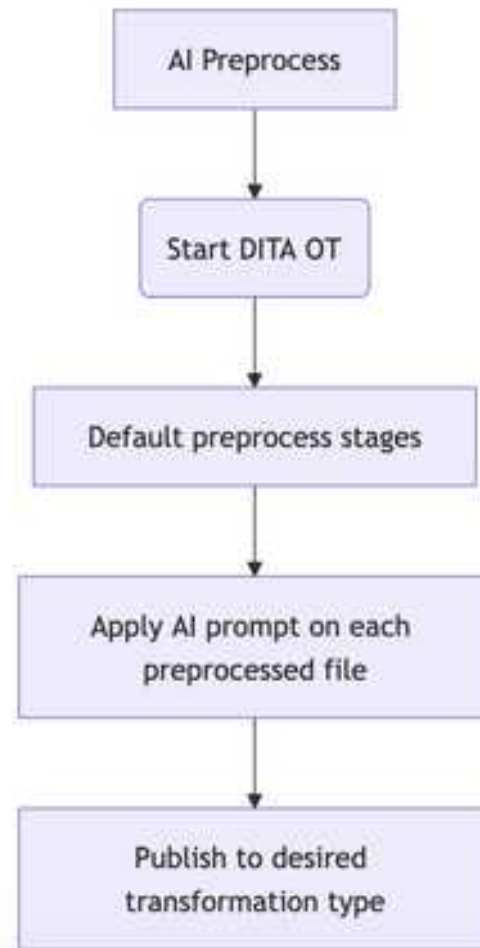
AI Preprocess Example – Translate at Publishing Time

(5) AI Summarize

- DITA OT Plugin which generates a summary of the entire publication, in the order specified in the DITA Map.
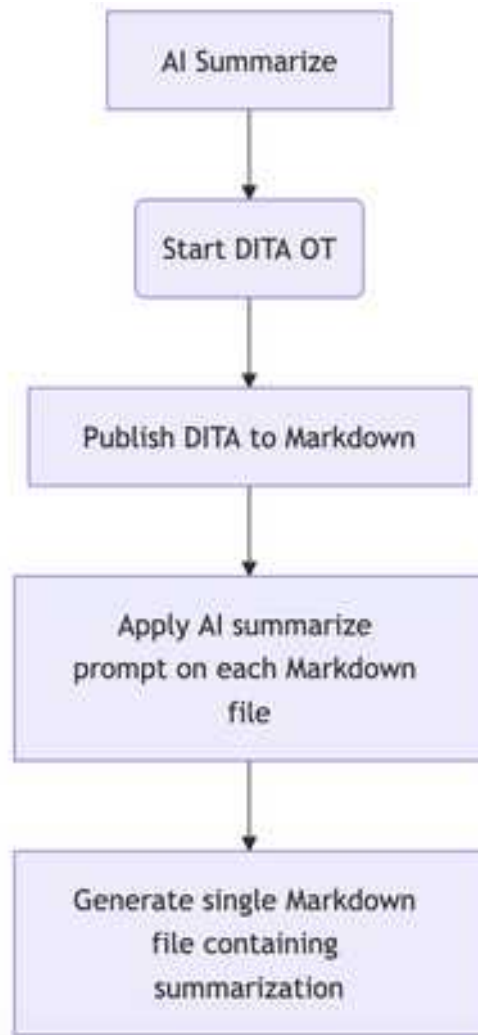
AI Summarize - Examples

- Suggest additional links to add between topics.

- Suggest place in DITA Map where to add a new DITA topic.

- Suggest ways to re-organize the user manual's DITA Map.

- Generate a graph showing relations between topics.

THANK YOU!

Questions time...

Radu Coravu

radu_coravu@oxygenxml.com