

Wiki <-> DocBook Conversion

FEUP - EdukaLibre

Wiki filter

Wiki Philosophy

- ◆ Wiki is intended to be as easy to read and write as possible but keeping the final document readability.
- ◆ Easy to learn and fast to write
- ◆ The syntax consists entirely of punctuation characters, such as asterisks, underscores, square brackets, etc...
- ◆ Markdown is a writing format while HTML is a publishing format

Wiki Standard

- ◆ There is no Wiki standard. Instead, there are several similar formats, but different enough to make it impossible to create a tool compatible with all formats

Main Wiki Styles

◆ Wiki Markup

- Large set of rules
- Widely used

◆ Wiki Markdown

- Punctuation characters carefully chosen so as to look like what they mean
- Easier to learn

Markdown

- ◆ “Punctuation characters have been carefully chosen so as to look like what they mean. E.g., asterisks around a word actually look like **emphasis**. Markdown lists look like, well, lists. Even blockquotes look like quoted passages of text, assuming you've ever used email”, John Gruber

Wiki Systems

- ◆ Are systems that help people to writing contents using the Wiki format, usually collaborative content
- ◆ These systems may have some features, such as: Authentication control; Version control; Edition control

Some Wiki Systems

- ◆ Wiki Wiki Web - <http://c2.com/cgi/wiki?WikiWikiWeb/>
- ◆ PhpWiki - <http://phpwiki.sourceforge.net/>
- ◆ Twiki - <http://twiki.org/>
- ◆ MediaWiki – engine used by Wikipedia
<http://sourceforge.net/projects/wikipedia/>

These systems use a format similar to Wiki Markup style

Wiki to Docbook Filter

- ◆ Wiki to Docbook filter is developed in PHP using the Text_Wiki package which is included in the Pear* framework.

* PEAR - PHP Extension and Application Repository

Text_Wiki

- ◆ Text_Wiki is an object-oriented tool which comes with its own set of markup rules.
- ◆ However, it allows setting any rules to match any style.
- ◆ The rules are configured through regular expressions

How Text_Wiki works

- ◆ Text_Wiki works in two stages: parse, and render

- ◆ Parse

When parsing, Text_Wiki makes a copy of the source text. It then applies "rules" (class files of **regular expressions** and processing commands) and looks for specific structured text (markup) in the source.

- ◆ Render

Rendering is the reverse of parsing.

Text_Wiki loads a format object (e.g., Docbook), and runs the pre-render method, after it runs the post-render method for the given format, and returns the newly formatted text.

Docbook to wiki conversion

- ◆ A new tool is being built from scratch, in PHP 5.
- ◆ Two native PHP function toolkits are being employed:
 - ◆ SimpleXML
 - ◆ XMLparser
- ◆ The structure of a Docbook document is considered in the development of the conversion algorithm.

The Docbook structure

```
<article>
  <articleinfo>
    <title>Edukalibre</title>

    <authorgroup>
      <author>
        <surname>Miguel Angel</surname>
        <firstname>Hernando Fernandez</firstname>
        <email>mhernand@gsync.escet.urjc.es</email>
      </author>
    </authorgroup>
  </articleinfo>

  <sect1>
    <title>Introduction</title>
    This would be <emphasis>italic</emphasis>.

    <sect2>
      <title>Another section</title>
      Some other text
    </sect2>
  </sect1>
</article>
```

Pure information content
Position is not relevant

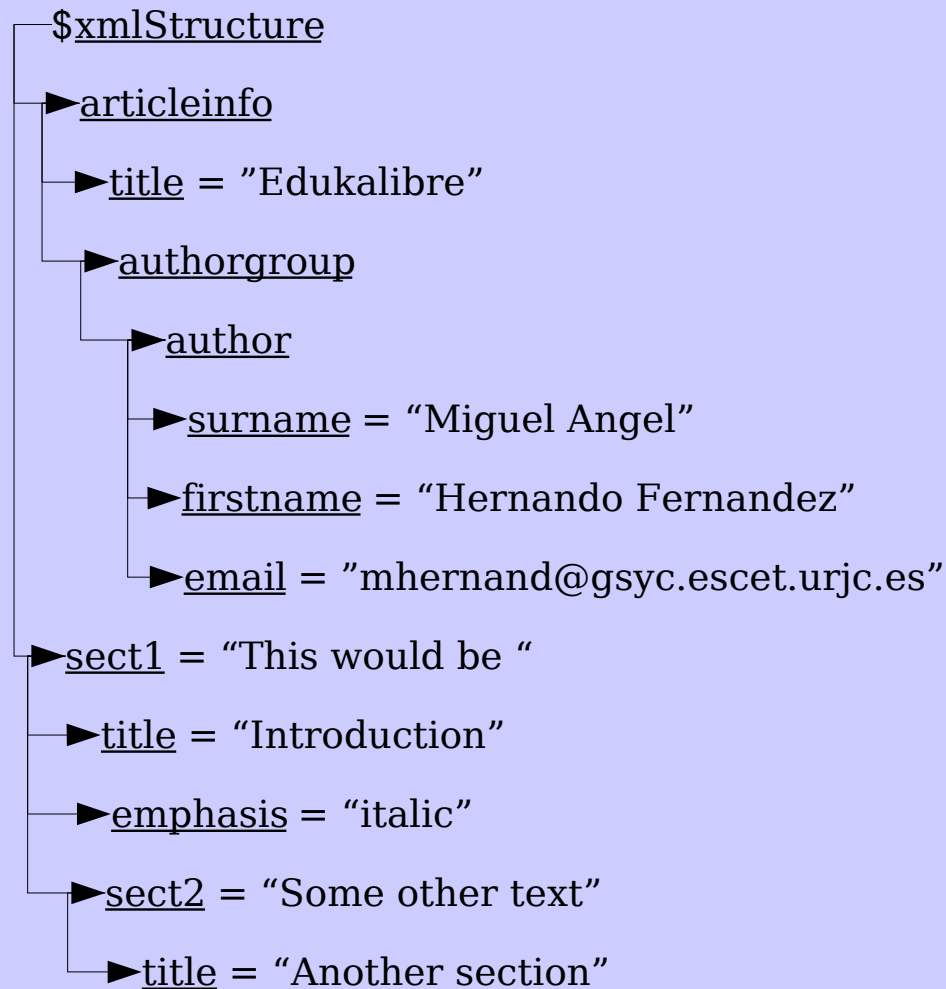
Relative information content
Position (within its scope) is
not relevant

Format information content
Position must be kept

Nested structures
Only relevant in relation
to information content

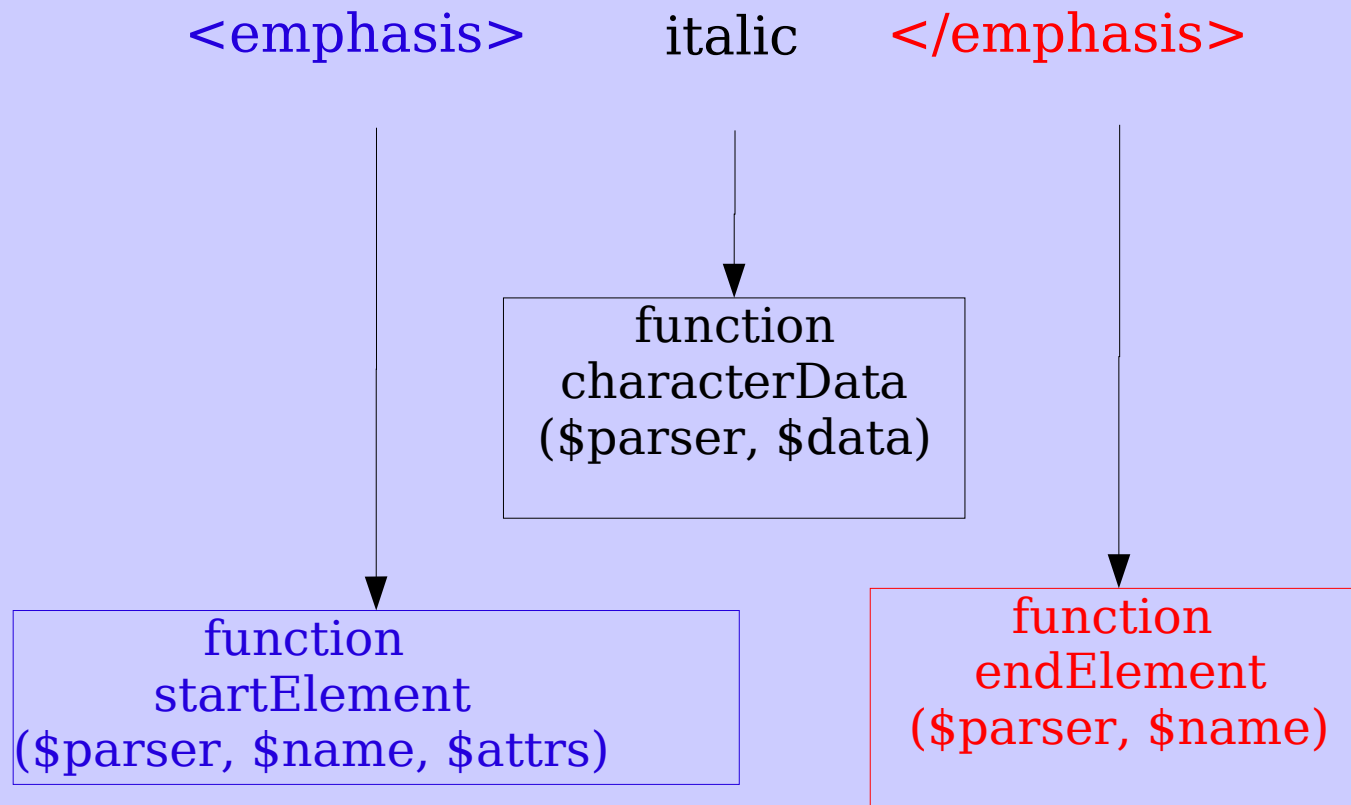
Two PHP XML tools (1)

Simple XML functions



Two PHP XML Tools (2)

XML parser functions



A Simple Algorithm

1 - Extract the document meta information

Document title, date and author data

2 - Start parsing the main body of the document

Each XML start or end tag is converted to the corresponding Wiki tag

3 - Whenever an element which may contain meta information appears, extract all its meta elements

i.e. whenever a section tag appears, extract its title

References

Markdown - <http://www.daringfireball.net/projects/markdown/>

Text_Wiki - http://wiki.ciaweb.net/yawiki/index.php?area=Text_Wiki

Pear - <http://pear.php.net/>

Docbook - <http://www.docbook.org/>

PHP - <http://www.php.net/>

FEUP Group members by alphabetic order

Guilherme Dutra

Jaime Villate

Nuno Faria