教科書アノテーションスキーマver.1

# 概要

本稿は、教科書のテキストデータの構造をアノテーションするためのスキーマである。

この文書に出てくるアノテーション例はすべて、東京書籍株式会社により提供された教科書でキスとデータを利用している。

# 文書情報

ルート要素は<textbook>である。source等以下の属性によって試験問題の基本情報を記述する。

<textbook>

@publisher　 発行元

@year　 年度

@subject　 科目名

# 目次

目次部分は<tableOfContents>でアノテーションする。

# 文書構造

# 部・編

教科書の内容は大きく分けて、「第○部」（社会系科目）、「第○編」（理科系科目）に分かれている。この部分は<part>でアノテーションする。

　<part>　部・編

# 章

各章は<chapter>でアノテーションする。

　<chapter>　章

# 節

章よりも細かい部分は一括して「節」とする。節は<section>でアノテーションする。

属性@levelを用いて、埋め込みのレベルを表示する。章よりすぐ下の節は@levelの値を1とし、埋め込みが深くなるにつれて値を大きくする。

<section> 節

@level 節の埋め込みのレベル（1以上の数字）

# 段落

各段落には<p>というタグをつける。

　<p> 段落

# 見出し

# 部・編の見出し

部・編の見出しは<partTitle>でアノテーションする。

<partTitle> 部・編の見出し

# 章の見出し

「第○章」で始まる見出しは<chapterTitle>でアノテーションする。

<chapterTitle> 章の見出し

# 節（番号付き）の見出し

章よりも細かく、かつ番号・記号が付いている部分（節）につけられた見出しは<sectionTitle>でアノテーションする。

<sectionTitle> 節の見出し

# 番号のつかない見出し

　番号のつかない見出しは、埋め込みの深さに関わらず<topic>でアノテーションする。

<topic> 番号のつかない見出し

# ページ番号

<pageNumber> ページ番号

@from ページ範囲の最初の番号

@to ページ範囲の最後の番号

（例）

複数ページからなる範囲が示されている場合：

<pageNumber from="11" to="12">p11-12</pageNumber>

単一のページの場合は、@fromと@toに同じ値を入れる。

<pageNumber from="20" to="20">p20</pageNumber>

# 改行

　<br> 改行位置を示す

# 図・表の情報

　<figure> 図の情報（図の番号および説明を含む）

@src 図のファイルへのリンク（あれば）

　<table> 表の情報（表番号および説明を含む）

@src 表のファイルへのリンク（あれば）