

# 「CD-毎日小学生新聞 記事データ集」データ構成

2017年2月  
日外アソシエーツ

## 1. ファイル形式

- ・ファイル形式は、tag形式 (“\tag\データ”の構成)、または、csv形式です。  
CD-ROM 1枚に tag形式, csv形式を併収、記事内容は同じです。
- ・データは原則として全角です (csv形式での記事全文の段落区切り;を除く)。

## 2. データ構成

- ・tag形式では、以下のtag構成・出現順です。
- ・csv形式では、各項目が” ”で囲まれたカラムで、以下の出現順です。

データ項目	tag	データ内容/備考
ID番号	ID	先頭‘S’+数字7桁の英数8桁 ★ユニーク
索引記事番号	C0	ex. S150101001 ★ユニーク
掲載面種別コード	AD	毎日小学生新聞は一律‘48’
写真, 図の有無	AE	有「Y」 or 無「N」
掲載日付, ページ	AF	ex.150101S01
著作権フラグ	ZZ	著作権の無い記事に‘著作権無’と表示
記事見出し	T1	
掲載日付, 掲載紙区別など	S1	日付/掲載紙/ページ/写真・図有無/文字数
記事前文	S2	記事の第一段落
記事全文	T2	
本文キーワード	KB	表記 (漢字)

## 3. データ内容

### 1) 著作権フラグ (tag.ZZ)

- ・tag形式では、著作権の無い記事にのみtagZZ「著作権無」を付与しています。
- ・csv形式では、著作権の無い記事に「著作権無」とデータ付与しています。

### 2) 記事見出し (tag.T1)

- ・著作権の無い記事にはデータ先頭に‘★’を付与しています。

### 3) 掲載紙 (tag.S1内)

- ・掲載日付の後の掲載紙は‘小学生新聞’‘小学生新聞別刷’‘こども新聞’‘学生’の4種類です。

### 4) 記事全文 (tag.T2)

- ・記事が複数の段落に分かれる場合、tag形式では段落ごとにtag.T2で区切り、csv形式では、同一カラム内で半角セミコロン (;) で区切っています。

## 4. 備考

- ・文字コードはShift-JIS、改行コードはCR+LFです。

以上