

# - ブログ出版局 MT互換 データ解説 -

# 1.この文書について

この文書はブログ出版局により本にすることができるMT互換データの解説です。MT互換データはMovableType(R)の出力するデータと互換性がありますが、**この文書はMovableType形式の正式な仕様書ではありません**。また、ブログ出版局はこの文書の正確性を保障しませんし、この文書を利用したことによる損害等に対して一切責任を負いません。

ブログ出版局はMovableTypeの開発元であるSix Apart社とは無関係です。この文書に関するコメントはdev [AT] cssj.jpまでメールしてください。

## 2.MT互換データとは

### 2.1 概要

MT互換データは、Six Apart社のMovableTypeによるブログデータのエクスポート形式と互換性を持ちます。MT互換データは、書き出したブログの全てのエントリで構成されており、各エントリについて、次の内容が含まれています。

- 本文
- 追記
- 読者から寄せられたコメント
- トラックバック

MT互換データはエントリの繰り返しであるため、ブログそのものについての情報(ブログのURL、タイトル、概要、管理者など)は含まれません。また、ブログに貼り付けられた画像や音声、Flashなどの情報も含まれません。

### 2.2 データ形式の詳細

MovableTypeによるブログデータのエクスポート形式は、[Six Apart社による解説](#)が公開されています。おおよそこの仕様に従って出力されたデータであれば、ブログ出版局により書籍化できます。この文書では、ブログ出版局のMT互換データ独自の拡張や、より詳細な仕様について記述しています。

## 3.ブログ出版局によるデータの認識

### 3.1 キャラクタ・エンコーディングと改行コード

MT互換データのキャラクタ・エンコーディングとして、UTF-8、シフトJIS、EUC-JPのいずれかを使用できます。しかし、MT互換データ自体には、そのデータがどのキャラクタ・エンコーディングを用いているかといった情報がないため、キャラクタ・エンコーディングを他の方法で明示する必要があります。

PHPのmbstring、Perlのjcode、JavaのJISAutoDetectといった機能を使って、どのキャラクタ・エンコーディングであるかを自動検出することも考えられますが、100%正確に検出する方法はありません。

シフトJIS、EUC-JPはいわゆる「機種依存文字」を含む非標準のコードを含むことができます。このようなコード体系には、PHPではSJIS-win,eucJP-win、JavaではMS932,EUC\_JP\_Solaris(JDK1.5.0以降)といった識別名が付けられています。

ブログ出版局ではUTF-8の使用を推奨します。PHPで内部エンコーディングとしてEUC-JPを用いている場合、直接UTF-8に変換すると機種依存文字が化けます。これを回避するためには、一旦データをSJIS-winへ変換し、さらにUTF-8へ変換してください。

改行コードはLFを推奨しますが、CRLF,CRも認識します。

### 3.2 サポートされるメタデータ

以下は、サポートされるメタデータの説明です。この一覧にないメタデータは無視されます。なお、現

在無視されるメタデータも将来的には有効になる可能性があります。

**表 3.1 メタデータの一覧**

名前	可能な値	デフォルト値	大文字小文字の区別	説明
AUTHOR	任意	-	あり	エントリあるいはコメントの投稿者の名前です。
CATEGORY	任意	-	あり	エントリが属するカテゴリです。
CONVERT BREAKS	0   __default__   default	0	-	0以外の場合、本文やコメントの改行を自動的に<BR;>タグに置き換えます。また、2つの連続する改行は段落(<P;>タグ)の境界に置き換えます。0の場合は何もしません。
DATE	説明を参照してください	当日の日付	なし	<p>エントリやコメントの投稿時刻を表します。日付の形式は以下の3つが許されます。なお、日付フォーマットの表記は <a href="#">JavaのSimpleDateFormatの説明</a> を参照してください。</p> <ul style="list-style-type: none"> <li>MM/dd/yyyy hh:mm:ss a</li> <li>MM/dd/yyyy HH:mm:ss</li> <li>yyyy/MM/dd HH:mm:ss</li> </ul> <p>以下は日付の表記の例です。いずれも2006年3月5日午後7時14分14秒を表します。</p> <p>03/05/2006 07:14:14 PM 03/05/2006 19:14:14 2006/03/05 19:14:14</p>
PRIMARY CATEGORY	任意	-	あり	
STATUS	Publish   Future   Draft	Publish	なし	

### 3.3 複数行フィールドのサポート

ブログ出版局は、複数行フィールドのBODY(本文),EXTENDED BODY(追記),COMMENT(コメント)を認識します。EXCERPT(概要),PING(トラックバック)は書籍には反映されません。

## 4.既知の問題

MT互換データは、MT互換データの構文要素がブログの内容に含まれる可能性を考慮していません。そのため、誤ってデータが認識されてしまう可能性があります。

### 4.1 区切り記号を含むデータについて

フィールドセパレータ(「-」の5文字だけからなる行)を本文やコメントの内容に含む場合、そこが複数行フィールドの終わりであると誤認します。

## 4.2 メタデータに類似した形式で始まる内容

コメントの内容がメタデータ(「名前:値」という形式の行)に類似している場合、その部分をメタデータと誤認します。