

## 「読売新聞記事データ集」TAG／データ構成

### 1. ファイルについて

- ・文字コードはShift\_JISです。
- ・DOS用改行コード（CR，LF）ではなく、UNIX用の改行コード（LFのみ）を使用しております。

### 2. データ形式

- ・形式は“\TAG\データ”
- ・TAG／データともにすべて全角ですが、一部、半角文字が含まれている場合があります。
- ・ファイルは月単位に分かれています。

### 3. TAG／データ構成

データ項目	TAG	データ内容／備考
ID番号	ID	00000010から10UP ★ユニーク
記事ID	C0	例) 20170612TYM01AA 001 ★ユニーク
頁	Y0	例) 01
面種コード	Y1	データ内容は下記参照
面名	Y2	例) 一面
版コード	Y3	例) 140
見出し(扱い)段数	Y4	例) 07
本文文字数	Y5	例) 1557
記事分類コード	Y6	データ内容は下記参照
写真・図表有無	YE	「写真」「表」「写真・表」の3パターン
掲載年月日	YF	例) 20170612
キーワード	KB	表記(漢字)
キーワード	AB	読み(カタカナ)
タイトル	T1	記事見出し
本文	T2	記事本文

### 4. データ内容

#### 1) 記事ID (TAG. C0)

- ・データ内容は以下のとおりです。

掲載年月日(8ケタ) + 紙誌名(3ケタ) + 頁(2ケタ) +

面種コード（2ケタ） + スペース（全角4文字） + ページ内記事通番（3ケタ）  
 例) \C0\20150109TYM11AG 002

2) キーワード (TAG. KB)

- ・2009年3月31日付までのデータには、記事の主題に見合った重要なキーワードの先頭に「\$」を付与しています（2009年4月1日以降は廃止）。

3) 写真・図表有無 (TAG. YE)

- ・写真や図表を含む記事にのみTAG/データを付与しています。

4) 紙誌名 (TAG. C0内)

- ・コード内容は以下のとおりです（中部支社は2002年に本社から支社に移行）。

T	O	S	C	Y	M	E
東京本社	大阪本社	西部本社	中部支社	読売新聞	朝刊	夕刊

例)

TYM	TYE	TMn	TE n
東京本社・読売新聞・朝刊	東京本社・読売新聞・夕刊	東京本社・朝刊・別刷n部	東京本社・夕刊・別刷n部
OYM	OYE	OMn	OE n
大阪本社・読売新聞・朝刊	大阪本社・読売新聞・夕刊	大阪本社・朝刊・別刷n部	大阪本社・夕刊・別刷n部
SYM	SYE	SMn	SE n
西部本社・読売新聞・朝刊	西部本社・読売新聞・夕刊	西部本社・朝刊・別刷n部	西部本社・夕刊・別刷n部
CYM	CYE	CMn	CE n
中部支社・読売新聞・朝刊	中部支社・読売新聞・夕刊	中部支社・朝刊・別刷n部	中部支社・夕刊・別刷n部

\* 中部支社には夕刊はありません。

\* n部が2なら朝刊別刷2部。

5) 面種コード (TAG. Y1) \*主なもの

AA	朝刊一面	UA	夕刊一面
AB	朝刊二面	UB	夕刊二面
AC	朝刊三面	UC	夕刊三面
AJ	朝刊2社面	UG	夕刊2社面
AK	朝刊社会面	UH	夕刊社会面

6) 版コード (TAG. Y3)

- ・朝夕刊の「版建て」に対応したコードが付与されています。

010~100=夕刊 110~200=朝刊

例えば夕刊の2版は「020」、朝刊の13版は「130」となります。

- ・3ケタ目にはKとSが付与されることがあります。

Kは選挙などで締切時間が延長された「改造版」のことです。

Sは朝刊と夕刊が配達される地域にのみ掲載される紙面（版）のことです。

## 7) 記事分類コード (TAG. Y6) 1987年1月～2023年9月

邦文記事分類	コード	詳細	邦文記事分類	コード	詳細
政治	Z01	政治	文化	V01	文化
	Z02	右翼・左翼		V02	学術
	Z03	選挙		V03	美術
	Z04	行政		V04	映像
	Z05	地方自治		V05	文学
	Z06	司法		V06	音楽
	Z07	警察		V07	演劇
	Z08	日本外交		V08	芸能
	Z09	軍事		V09	舞踊
	Z10	戦争		V10	宗教
経済	Y01	経済	生活	U01	生活
	Y02	財政		U02	健康
	Y03	金融		U03	衣
	Y04	企業		U04	食
	Y05	中小企業		U05	住
	Y06	技術		U06	余暇
	Y07	情報		U07	行事
	Y08	サービス	犯罪・事件	T01	犯罪・事件
	Y09	貿易		T02	事故
	Y10	国土・都市計画		T03	災害
	Y11	鉱工業	科学	S01	科学
	Y12	資源・エネルギー		S02	宇宙
	Y13	農林水産		S03	地球
社会	X01	社会	S04	理工学	
	X02	市民運動	S05	生命科学	
	X03	社会保障	S06	動植物	
	X04	環境	国際	R01	国際
	X05	婦人		R02	アジア・太平洋
	X06	子供		R03	南北アメリカ
	X07	中高年		R04	西欧
	X08	勲章・賞		R05	旧ソ連・東欧
	X09	労働		R06	中東
	X10	教育		R07	アフリカ
スポーツ	W01	スポーツ	皇室	Q01	皇室
	W02	巨人軍			

## 7) 記事分類コード (TAG. Y6) 2023年10月～

邦文記事分類	コード	詳細	邦文記事分類	コード	詳細		
政治	Z01	政治	文化	V01	文化		
		右翼・左翼			学術		
		選挙			美術		
		行政			映像		
		地方自治			文学		
		司法			音楽		
		警察			演劇		
		日本外交			芸能		
		軍事			舞踊		
		戦争			宗教		
経済	Y01	経済	生活	U01	生活		
		財政			健康		
		金融			衣		
		企業			食		
		中小企業			住		
		技術			余暇		
		情報			行事		
		サービス			犯罪・事件	T01	犯罪・事件
		貿易	事故				
		国土・都市計画	災害				
		鉱工業	科学	S01			科学
		資源・エネルギー					宇宙
		農林水産			地球		
社会	X01	社会	国際	R01	国際		
		市民運動			アジア・太平洋		
		社会保障			南北アメリカ		
		環境	西欧				
		婦人	旧ソ連・東欧				
		子供	中東				
		中高年	アフリカ				
		勲章・賞	皇室	Q01	皇室		
		労働					
教育							
スポーツ	W01	スポーツ					
		巨人軍					