

「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. CO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	Z 0 1	政治	文化	V 0 1	文化
	Z 0 2	右翼・左翼		V 0 2	学術
	Z 0 3	選挙		V 0 3	美術
	Z 0 4	行政		V 0 4	映像
	Z 0 5	地方自治		V 0 5	文学
	Z 0 6	司法		V 0 6	音楽
	Z 0 7	警察		V 0 7	演劇
	Z 0 8	日本外交		V 0 8	芸能
	Z 0 9	軍事		V 0 9	舞踊
	Z 1 0	戦争		V 1 0	宗教
経済	Y 0 1	経済	生活	U 0 1	生活
	Y 0 2	財政		U 0 2	健康
	Y 0 3	金融		U 0 3	衣
	Y 0 4	企業		U 0 4	食
	Y 0 5	中小企業		U 0 5	住
	Y 0 6	技術		U 0 6	余暇
	Y 0 7	情報		U 0 7	行事
	Y 0 8	サービス	犯罪・事件	T 0 1	犯罪・事件
	Y 0 9	貿易		T 0 2	事故
	Y 1 0	国土・都市計画		T 0 3	災害
	Y 1 1	鉱工業	科学	S 0 1	科学
	Y 1 2	資源・エネルギー		S 0 2	宇宙
	Y 1 3	農林水産		S 0 3	地球
社会	X 0 1	社会		S 0 4	理工学
	X 0 2	市民運動		S 0 5	生命科学
	X 0 3	社会保障		S 0 6	動植物
	X 0 4	環境	国際	R 0 1	国際
	X 0 5	婦人		R 0 2	アジア・太平洋
	X 0 6	子供		R 0 3	南北アメリカ
	X 0 7	中高年		R 0 4	西欧
	X 0 8	勲章・賞		R 0 5	旧ソ連・東欧
	X 0 9	労働		R 0 6	中東
	X 1 0	教育		R 0 7	アフリカ
スポーツ	W 0 1	スポーツ	皇室	Q 0 1	皇室
	W 0 2	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	Z 0 1	政治	文化	V 0 1	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	Y 0 1	経済	生活	U 0 1	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	T 0 1	犯罪・事件
		貿易			事故
		国土・都市計画			災害
		鉱工業	科学	S 0 1	科学
		資源・エネルギー			宇宙
		農林水産			地球
社会	X 0 1	社会			理工学
		市民運動			生命科学
		社会保障			動植物
		環境	国際	R 0 1	国際
		婦人			アジア・太平洋
		子供			南北アメリカ
		中高年			西欧
		勲章・賞			旧ソ連・東欧
		労働			中東
		教育			アフリカ
スポーツ	W 0 1	スポーツ	皇室	Q 0 1	皇室
		巨人軍			