

多様なテキストデータの利用事例

プレーンテキスト、Word、一太郎、PDFなど、テキストデータには 多くの形式があります。入稿したテキストを様々な方法で加工し 効率的にDTP組版を行っている事例を紹介します。

雑誌書籍、その他あらゆる印刷物において、文字情報は最も基本的な構成要素の一つです。当社の創業当時は、原稿用紙に手書きされた文字原稿をお客様から受け取るのが当たり前でしたが、現在では文字情報はテキストデータとして入稿いただくのが一般的です。

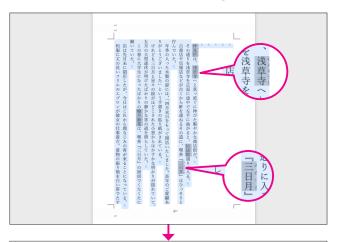
テキストデータは作成環境によって多くの形式が存在しており、その中には文章情報だけでなく書体や色などの文字スタイルや、ルビ等の組指定の情報が含まれている場合もあります。

当社は文字情報を扱うプロフェッショナルとして、 テキストデータをより効率的、かつ作成者の意図通り にDTP組版に反映させる方法を常に追求しています。 そのような事例をいくつかご紹介します。

事例1 一太郎で支給されたテキストデータの処理

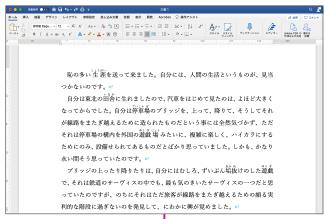
日本語ワープロソフトとして一定の人気を持つ一太郎。特に教育書などの原稿制作に現在も広く使用されているようです。

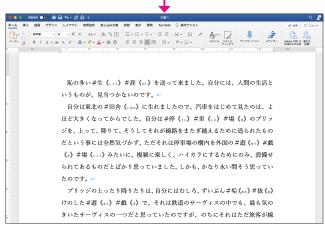
しかし、一太郎のデータはInDesignに取り込むこと



WordデータをInDesignに配置後。1行目の浅草寺の「浅」や、3行目の 三日月の「三」などが飛んでしまっています。 ができません。そのため、一太郎のデータを一旦 Word形式に変換してからInDesignに取り込むという 方法が用いられますが、一太郎のルビ機能が使われて いる場合、Wordに変換した時点では問題がなくても、 InDesignに取り込んだ際にルビの親文字が欠落するな どのトラブルが起きることが報告されています。

当社では、まずWord形式に変換した段階でマクロを使ってルビをタグ付きテキストに変換、それをInDesignに取り込んでから、スクリプトを使ってタグをルビに戻すという処理を行うことでトラブルを回避しました。また、一太郎のルビがすべてグループルビになっていたため、当社のルビ入力支援ソフトを使い、モノルビへの変換も行っています。





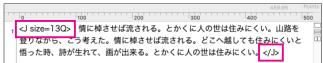
Wordのマクロを用いてルビをタグ付きテキストに変換。それをInDesign に取り込んでからスクリプトでルビに変換した。



事例2 MC-B2から書き出されたタグ付きテキスト

MC-B2で組版された漢文の学習書をInDesignで組み直すという改訂版制作の依頼がありました。MC-B2とInDesignの間にデータの互換性はありません。そのため、MC-B2から書き出されたテキストをInDesignのフォーマットに流し直さなければいけませんでした。しかし、MC-B2のテキスト構造は特殊で、文字のスタイル情報がすべてタグとして定義されています。さらに内容が漢文ということもあり、その量も膨大でした。そこで、このタグをInDesignのスタイルに置き換えることにしました。

まず、すべてのMC-B2タグについてInDesignのどのスタイルと紐付ければよいのか調査しました。そして必要な100種類ほどの検索置換設定を作成し、それらを一括して実行するスクリプトを開発。作業効率化と同時にヒューマンエラーの防止も実現できました。



上図の囲み部分がMC-B2のタグ。これを検索しInDesignのスタイル に一括して置き換えることで効率化を図った。

事例3 ルビだけが記号で括られたテキストデータ

ルビ入り書籍の文字原稿は様々な形式で入稿されます。最近ではWordのルビ機能が使われることが多くなりましたが、下図のようにルビ文字だけが記号類で括られた状態で入稿されたケースがありました。このままではどの文字に対してルビを付与するのかコンピュータには判断することができません。

そこで、ルビ文字が括られている記号の前にある1 文字以上の漢字を正規表現を用いて検索し、ヒットした文字列にルビを付与するスクリプトを開発しました。正規表現とは文字列中の「パターン」を検索するための記述方法です。ルビが非常に多い書籍だったため大きな効率化となりました。

がんじ漢字

のルビ入りテキスト

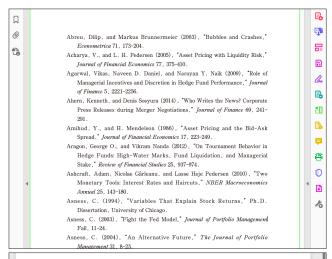
望ましい状態 # 漢 | かん | # 字 | じ |

原稿状態 漢字 | かんじ |

事例4 PDFのイタリック体を再現する

他書籍の紙面を流用する作業で、流用元データがPDFしかないというケースがありました。幸いテキストはアウトライン化されておらず抽出できる状態でしたが、単純にテキストとして取り出してしまうと、原書のイタリック体のスタイル情報が飛んでしまい、InDesignに流し込んでから再度手動で設定し直さなければいけません。これでは効率も悪くミスが起きる恐れもあります。

そこでまずPDFデータを、Acrobatを使用してイタリック体が再現できるWord形式へ変換しました。次にWord上でイタリック体に適当な文字色を設定してInDesignに流し込みます。InDesignで文字色を検索しイタリック体に置き換えることで対処できました。



Abreu, Dilip, and Markus Brunnermeier (2003), "Bubbles and Crashes,"

Econometrics 71, 173-204."

Acerya., and L. H. Pedersen (2005), "Asset Pricing with Liquidity Risk,"

Journal of Financial Economics 77, 375-410.9

Agarwal, Vikas, Naveen D. Daniel, and Narayan Y. Naik (2009), "Role of Managerial incentives and Discretion in Hedge Fund Ferformance," Journal of Finance 5, 2221-2256.6

Ahern, Kenneth, and Denis Sosyura (2014), "Who Writes the News? Corporate Press Releases during Merger Negotiations," Journal of Finance 69, 241-231.*

Amily And H. Mendelson (1986), "Asset Pricing and the Bid-Ask Journal of Financial Economics 17, 223-249.6

Aragon George 0, and Vikram Anada (2012), "On Tournament Behavior in Hedge Funds: High-Water Marks, Fund Liquidation, and Managerial Stake," Review of Financial Studies 25, 397-397-4.6

Asheraft, Adam, Nicolae Garleanu, and Lasse Heje Pedersen (2010), "Two Monetary Tools: Interest Rates and Haircuts, "NEBR Macroeconomics Annual 25, 143-180.6

Asness, C. (1904), "Variables That Explain Stock Returns," Ph.D. Dissertation, University of Chicago. "Asness, C. (2007), "Fight the Fed Model, "Journal of Portfolio Management Fall, 11-24.6

Asness, C. (2007), "Box Became a Quant, "How I Became a Quant, Insights from 25 of Wall Street's Effice Richard R. Lindsey and Barry Schachter (Eds.), John Wiley and Sons, Hoboken, NJ.

Asness, C. (2007), "How I Became a Quant, "How Became a Quant, Insights from 25 of Wall Street's Effice Richard R. Lindsey and Barry Schachter (Eds.), John Wiley and Sons, Hoboken, NJ.

Asness, C. (2007), "How I Became a Quant, "How Hexame a Quant, Insights from 25 of Wall Street's Effice Richard R. Lindsey and Barry Schachter (Eds.), John Wiley and Sons, Hoboken, NJ.

Asness, C. 4, Prazzini, and L. H. Pedersen (2013), "Quality Minus Junk," Working paper, AQR Capital Management and New York University."

Asness, C. 4, Frazzini, and L. H. Pedersen (2013), "Quality Minus Junk," Working paper, AQR Capital Management and New York University."

Asness, C. 4, Frazzini,

上がPDF、下がPDFから書き出したWordデータ。Wordデータ上で 斜体の部分に色をつけてInDesignに取り込むことで斜体を再現。

テキストデータの入稿方法等について、もしお困り のことがありましたら是非ご相談ください。最も効率 的で、ミスの少ない利用方法を提案いたします。