

新提案!!

オレンジの PDF & Excel でできる索引抽出・読み入力法

2023.5.25 作成

使用ソフト **Foxit PDF Reader** (フリーソフト), **Microsoft Excel**

新型コロナウイルス感染症のパンデミック以降、出版社も印刷会社も、校正の往来にも PDF を利用する機会が格段に増えました。

永和印刷 (以下、弊社) では、Prog 37 で Adobe Acrobat やいきなり PDF といった専用のソフトがなくても、PDF の結合・分割・圧縮・編集、Office 系ソフトや JPEG との相互変換などが無料でできるオンラインツール「I LOVE PDF」(<https://www.ilovepdf.com/ja>) をご紹介しました。今回は **Foxit PDF Reader (オレンジの PDF)** で PDF から索引用語等を抽出し、**Excel** で読みを自動入力できる方法を提案します。これを使えば、抽出用語を紙に転記したり、抽出用語を Excel にコピーし、読みを入力する必要もありません。従来のように印刷会社が抽出したリストを待ってから確認する手間と時間が大幅に軽減されます!! ちなみに、弊社へ永和印刷はハイライトを付けた PDF を御支給いただければ OK です!!

PART 1 Foxit PDF Reader でハイライトの用語を一気に抽出する!!

STEP 1 Foxit PDF Reader をダウンロードする。

ブラウザで <https://www.foxit.co.jp/downloads/> にアクセスし、ダウンロードをクリック。



Foxit PDF Reader とは?

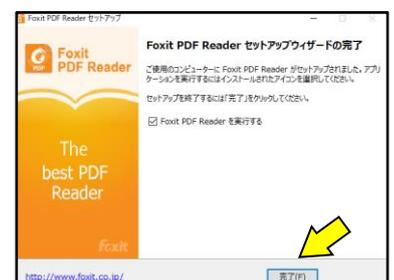
Windows 10/11 対応の (株) FoxitJapan 製のコンパクトで速度性能が高く、機能豊富なフリーの PDF ビューワー。「時代は赤からオレンジへ!」というキャッチが示すように、オレンジ色を基調としたデザインが特徴的です。

フリーソフトとはいえ、十二分な機能を備えています。

窓の杜 (<https://forest.watch.impress.co.jp/library/software/foxitreader/>) でもダウンロード可能です。



→ ダウンロードしたファイル (foxit-reader_12.1.0.15250_jp.exe) をクリックしてセットアップウィザードを開始しインストール完了してください。



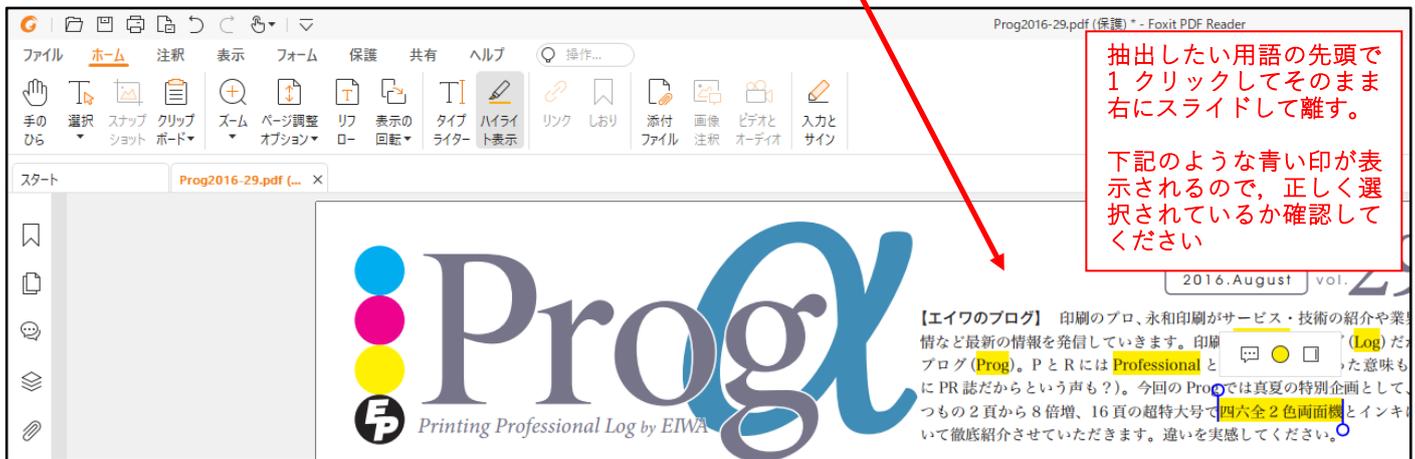
注：インストールおよびご利用については自己責任でお願いいたします。

STEP 2 PDF を準備する（印刷会社等に 1 ファイルでのノンブルの通った PDF の作成をご依頼ください）。

Point ! 単語だけではなく、ノンブルも抽出できます。ただし、ノンブルと PDF の頁番号を一致させておく必要があるため、ずれないように白頁も入れておきましょう（Reader ではできない作業です）

STEP 3 抽出用語をハイライト表示にしていく（この作業は Adobe Reader でも可能です）。

「ホーム」⇒「ハイライト表示」をクリックし、抽出したい用語を選択していきます。



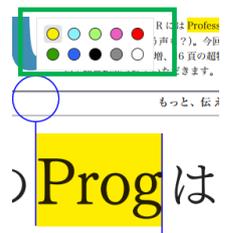
Point ! 「エリアハイライト」ではなく、「テキスト」をハイライトにしてください。

「検索してハイライト」から、 にチェックをつけてハイライトにすることも可能です。



Point ! 複数で作業する場合は、ハイライトの色分けをしておくとも便利です。

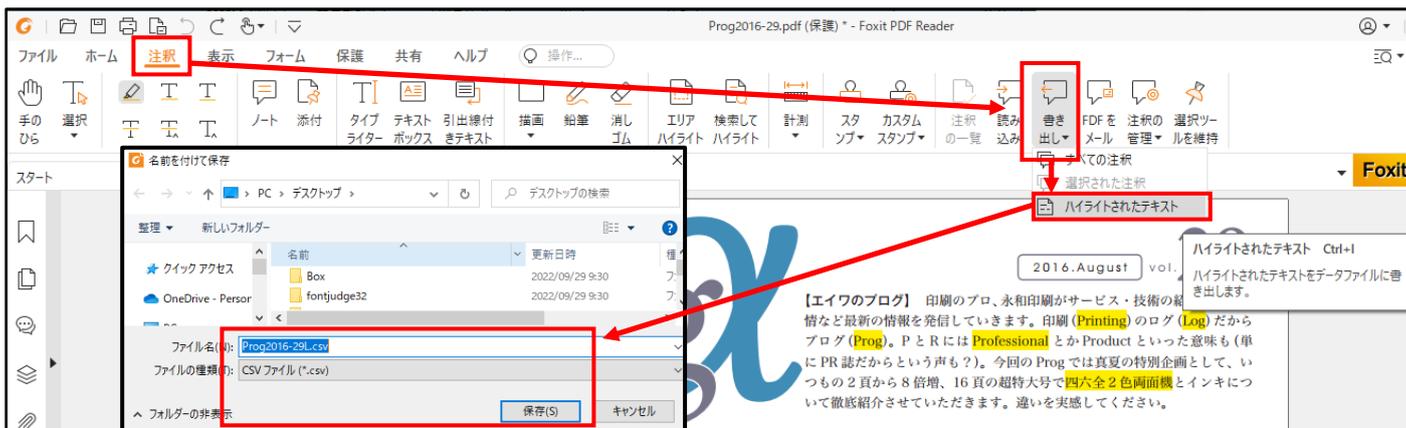
ただし、色分けは抽出データには反映できません。



Prog は

STEP 4 ハイライト表示のチェックの済んだ PDF を Foxit PDF Reader で開く (Adobe Reader では不可)。

「注釈」⇒「書き出し」⇒「ハイライトしたテキスト」を選択し、名前を付けて「保存」
(CSV ファイルか TXT 形式ファイルが選べます)

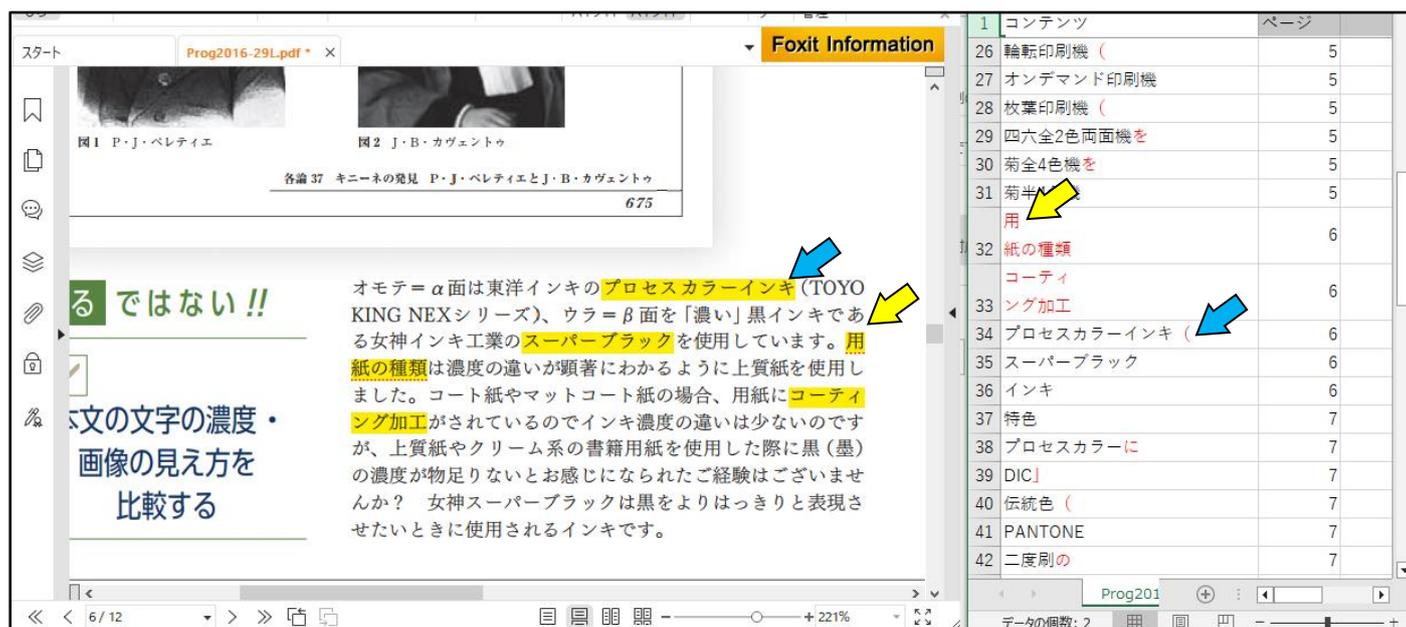


STEP 5 Excel で CSV ファイル (もしくは TXT ファイル) を開き、改行や余分な文字等を削除する。

※単語と頁が抽出されますが、改行や選択していない余分な文字・空白が入ることがあります。

Foxit PDF Reader の画面

Excel の画面



STEP 5 ではひと手間かかってしまうのが難点ですが、コピペを連続していくよりも早く確実ですので、興味がある方はぜひお試しください。実験用のデータとして、永和印刷のホームページから上記の Prog 29 のマーカー付 PDF がダウンロードできます (<http://www.eiwa-p.com/04progimage/Prog2016-29.pdf>)。

この段階では「読み仮名」が入っていません。読み仮名をご自身で Excel を使って自動で入れたい方は PART 2 へお進みください。読みやソートを制作会社・印刷会社に依頼する方は、データをお渡してください。

※Foxit PDF Reader にはそれ以外にも便利な機能がたくさんあります。PART 3 (p.7~) をご覧ください。

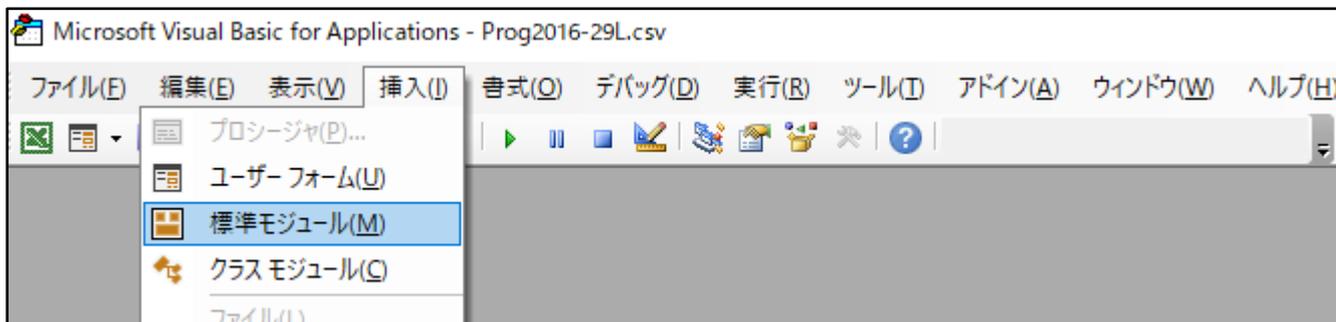
PART 2 Excelで抽出語の読み(ふりがな)を一気に書き出す!!

STEP 1 PART 1 で作成した CSV ファイルを Excel で開く。

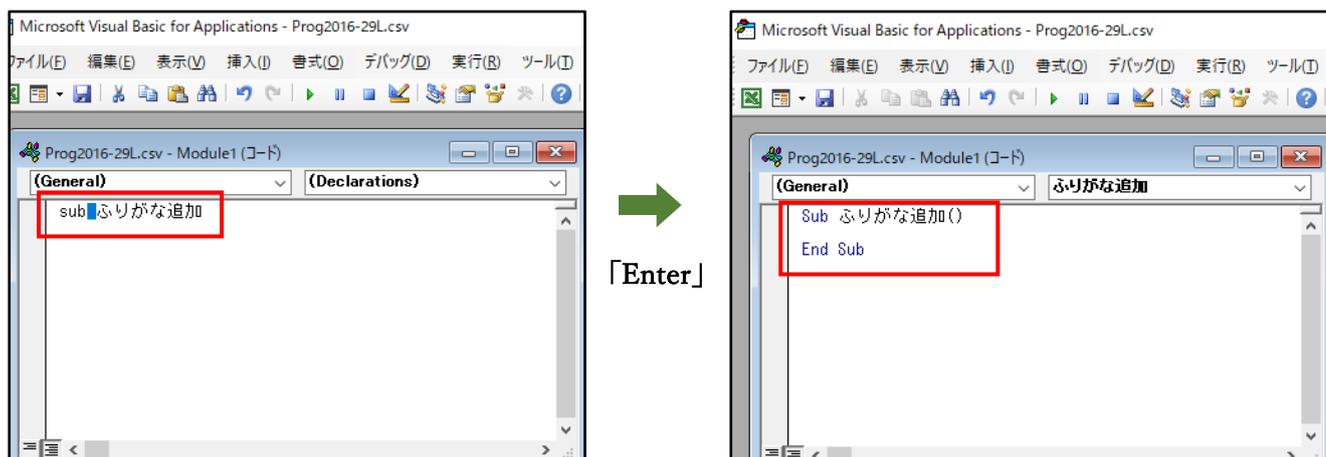
※単語と頁が抽出されていますが、改行や余分な文字が入っていることがあるので確認しておく。

(→PART 1 STEP 5 参照)

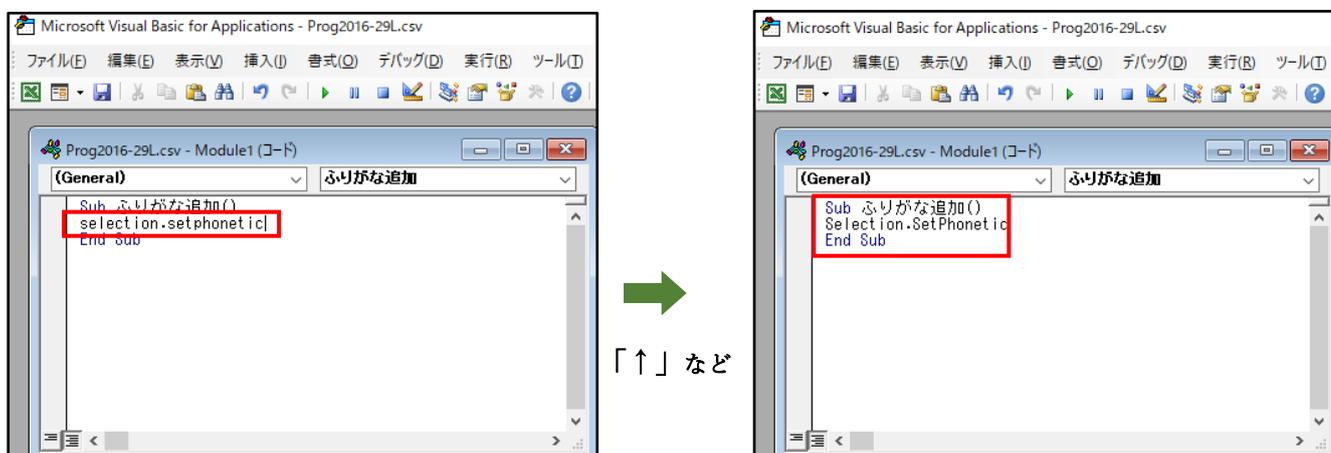
STEP 2 「Alt」 + 「F11」 で、編集画面を開き、「挿入」 → 「標準モジュール」 をクリック。



STEP 3 「sub ふりがな追加」 (sub は半角小文字で、sub と「ふ」の間は半角スペース) と入力し「Enter」。
→先頭の「s」が大文字になり、括弧「()」が追加されます。



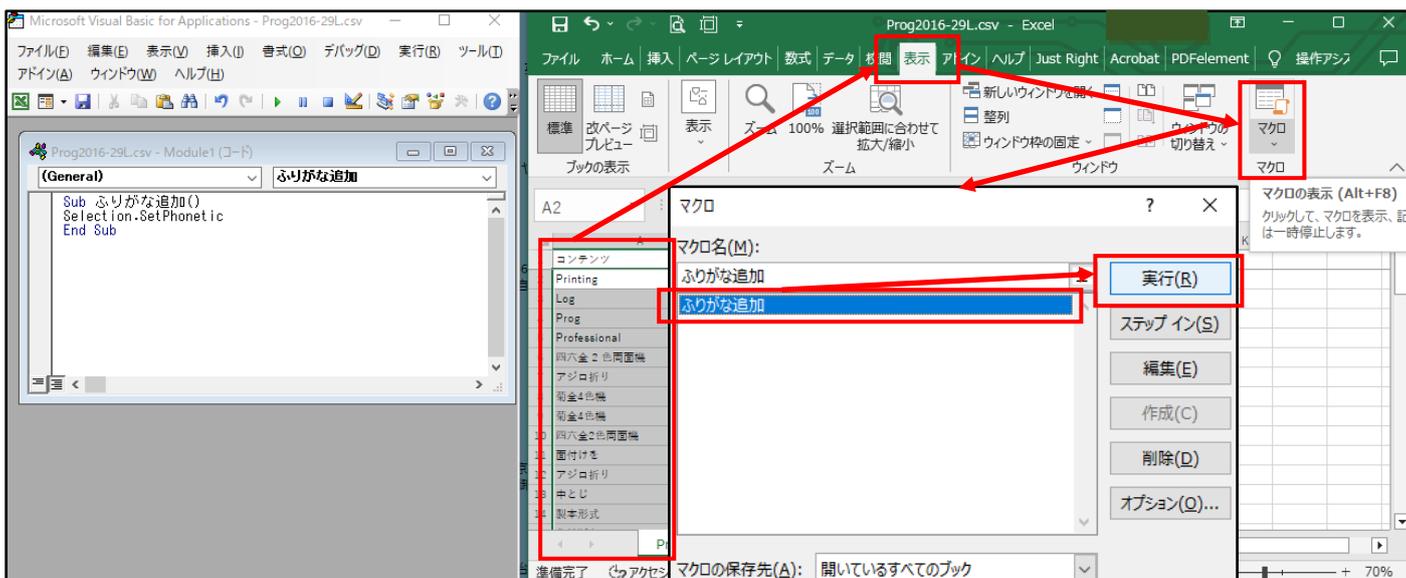
STEP 4 「Sub ふりがな追加 ()」と「End Sub」の間の行に「selection.setphonetic」と入力(↑などを押す)。
→先頭の「s」が大文字になります。



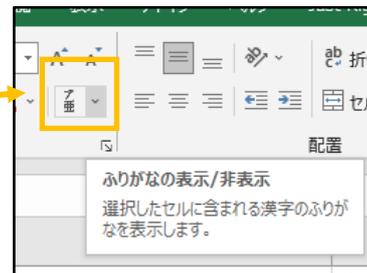
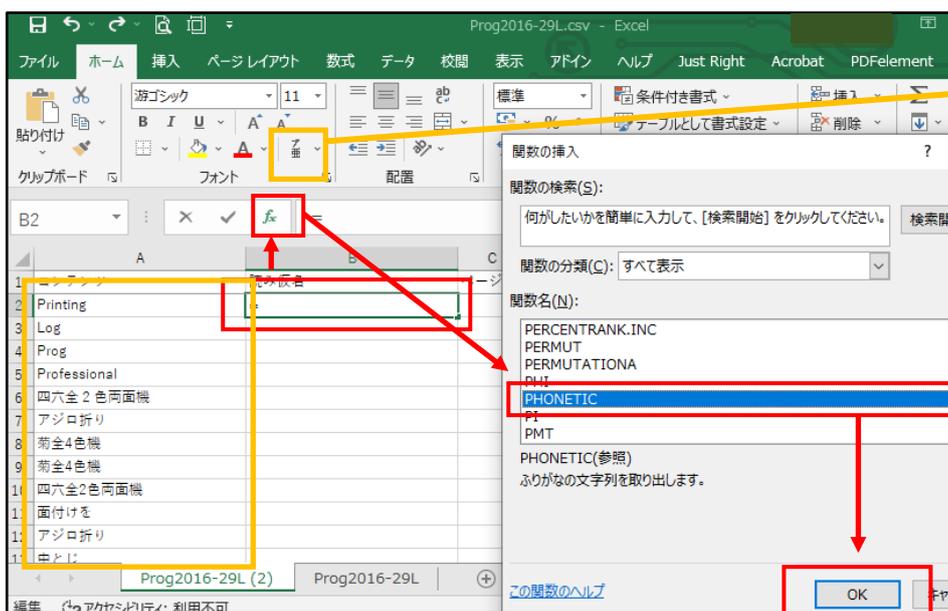
STEP 2~4 は VBA の作業で不安に感じるかもしれませんが、設定は以上です。挑戦してみてください!!
作業前には必ずバックアップを取っておいてください

STEP 5 「Alt」 + 「F11」 を押し、Excel 画面に戻る。

読みを追加したいセル範囲を選択し、[表示]→「マクロ」をクリックし、マクロ名「ふりがな追加」を選択して「実行」をクリック → これでデータ的にはふりがなが追加されている状態になります。

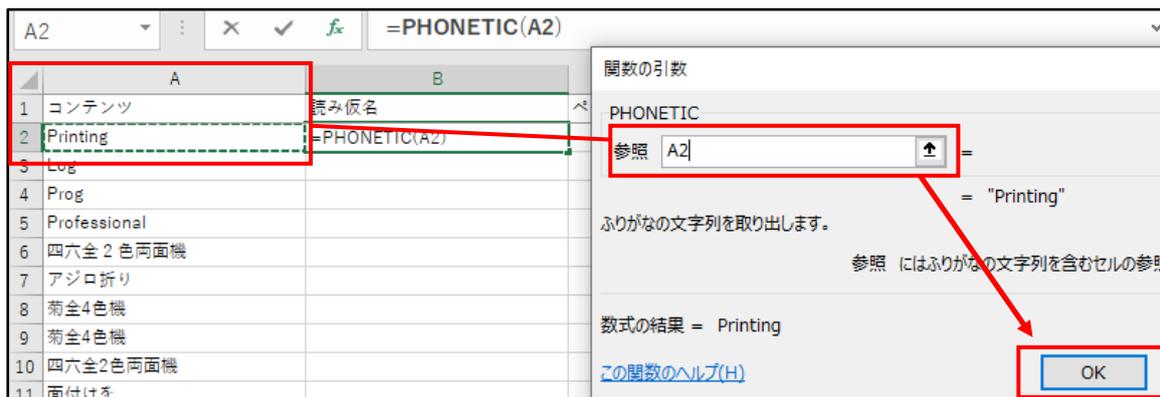


STEP 6 タイトル行「コンテンツ」の右側 (B 列) に列を挿入 (「読み仮名」と入力) し、2 行目以降に PHONETIC 関数 (ふりがな) を使って、ふりがなを取り出す。関数の挿入「fx」→「PHONETIC」を選択し「OK」をクリック。

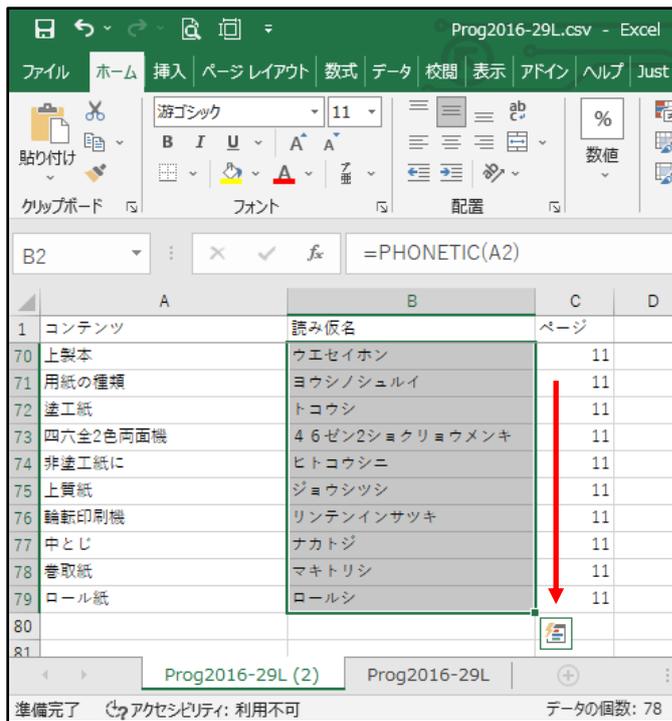


Point ! 「ふりがなの表示」をクリックするとふりがなが表示されます。「ふりがなの設定」でひらがな・カタカナの選択が可能です。

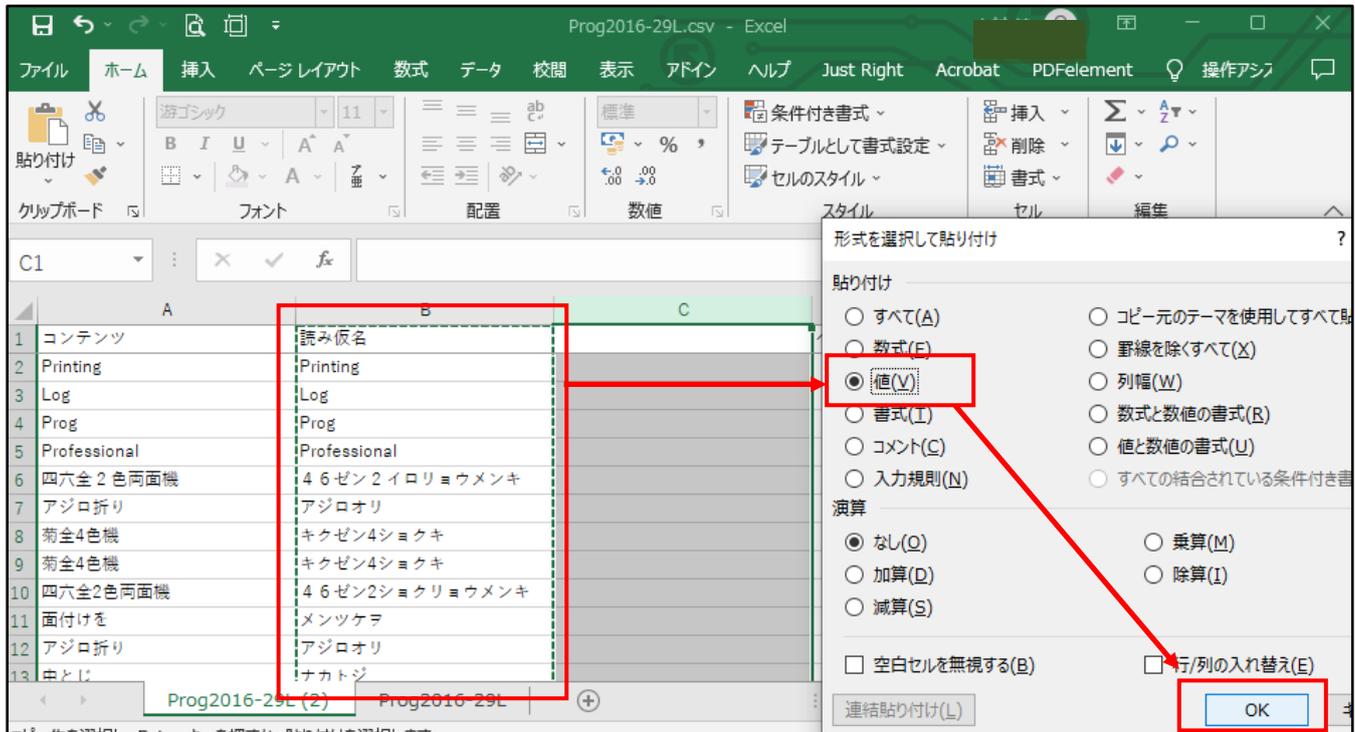
→ 参照に左隣の「A2」(対象の文字列) を選択し「OK」をクリックすると、ふりがなが反映されます。



STEP 7 「B2」セルの右下にカーソルを合わせて、下にドラッグし、ふりがなを表示させる。



STEP 8 このままだとセル内が数式になっていて修正ができないので、列を挿入して、「B 列」をコピーしたあと、「形式を選択して貼り付け」⇒「値」を選択して、「OK」で、ふりがなをテキスト化する。
⇒ふりがなを確認して、適切でないものは修正してください。



Point! 「ふりがな」を修正することも可能ですが、作業的には「値」でテキストにしたもの（C列）を直すほうが簡単で効率的なのでオススメです。

⇒単純 50 音であれば、「データ」⇒「フィルター」で並び替えれば、索引の基本データの完成です。
おつかれさまでした。

PART 3 Foxit PDF Reader の利便性の検証 (Acrobat Reader と徹底比較)

Foxit PDF Reader (以下, Foxit) の特徴は, コンパクトで速度性能が高く, 機能が豊富なことが挙げられます。使い勝手もよいので, Acrobat Reader (以下, Adobe) と比較してみました!! 詳しい使い方はマニュアル (https://www.foxit.co.jp/wp-content/uploads/FoxitPDFReader12.0_QuickGuide.pdf) をご覧ください。

比較1 テキストのスタイルの設定

共通項目=フォント, ポイント, 色, ボールド, イタ, 下線, 取消線, 上付, 下付, 左右中央両端揃え
Foxit あり, Adobe なし = 箇条書き, 段落番号, 行間, 文字幅, 文字間の設定

Foxit PDF Reader



Acrobat Reader

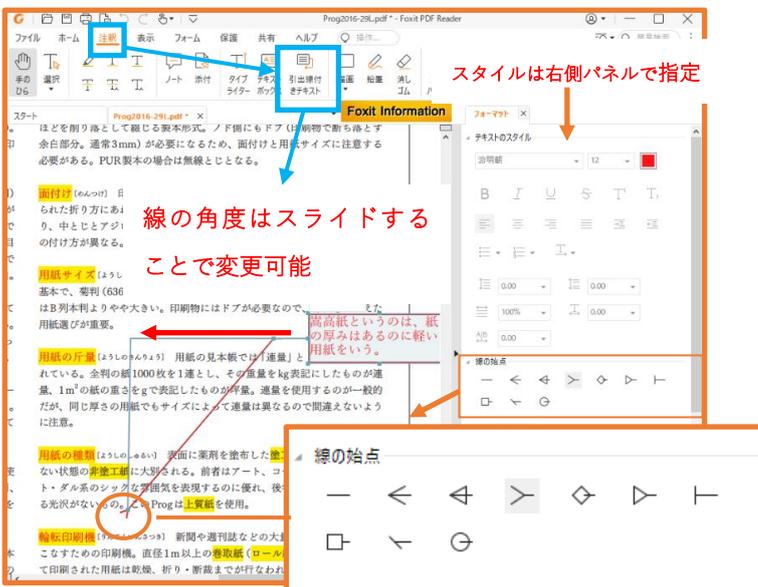


Foxit でも, Acrobat でも, ポイントを変更すれば行間はなりゆきで変更されますが, スペースの都合等があるときには, Foxit の細かく設定できる機能はととても便利です。

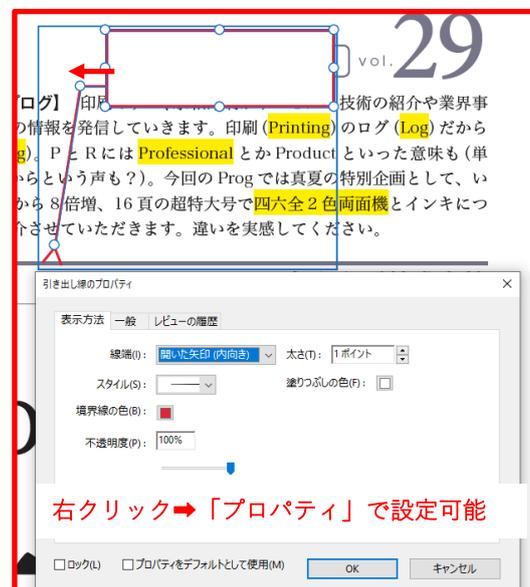
以下の機能も同様ですが, Foxit は Word や Excel などと同様のリボン表示と右側のパネルで指定が可能です, Acrobat はそれぞれの動作を右クリックから選択 or 右クリック⇒プロパティから選択の形なので, 操作性は Foxit のほうが優れています。

比較2 引出線付テキストの機能 「引出線付テキスト」を選択し, 開始点を決定し, ドラッグしてボックスを作成, テキストを入力。この機能を使えば, 別途, 線を挿入する手間が省けます。

Foxit PDF Reader

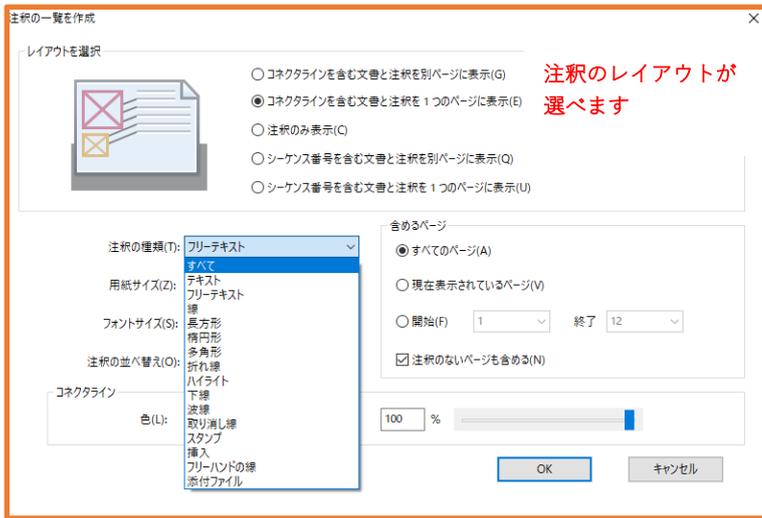


Acrobat Reader

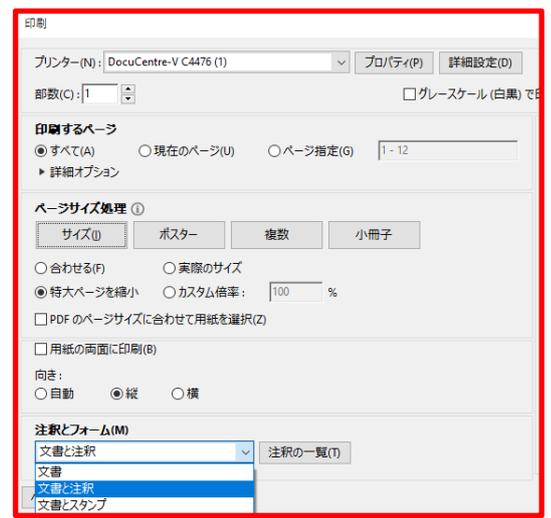


Acrobat では「表示」⇒「ツール」⇒コメント
⇒右の…⇒「引出線付きテキストボックス」

Foxtit PDF Reader



Acrobat Reader

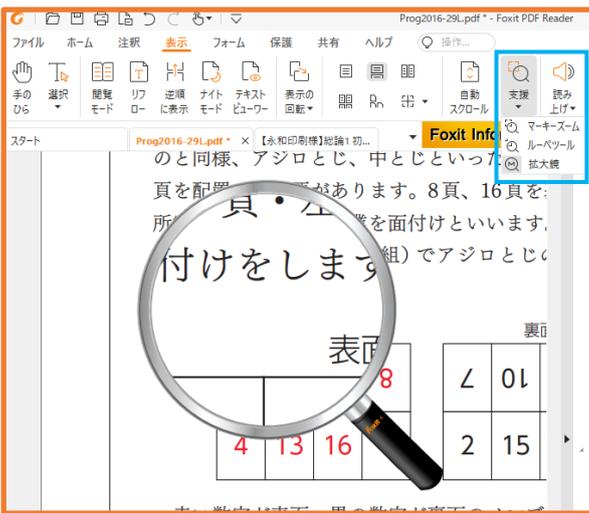


「注釈」⇒「注釈の一覧」で、コネクタ付きの注釈や注釈のみ表示、番号付き表示、見開き、別頁、注釈の種類などの詳細な設定が可能

「印刷」の設定で注釈の表示の設定が可能

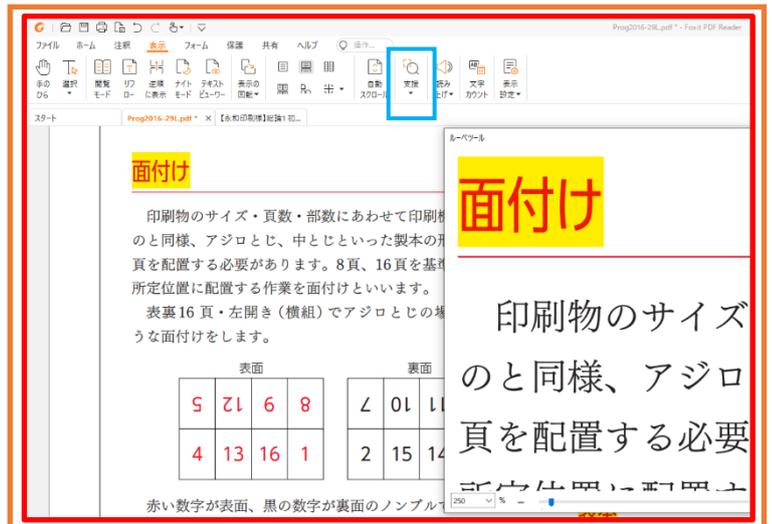
比較 5 拡大表示機能

Foxtit PDF Reader



「表示」⇒「支援」⇒「拡大鏡」

Foxtit PDF Reader / Acrobat Reader

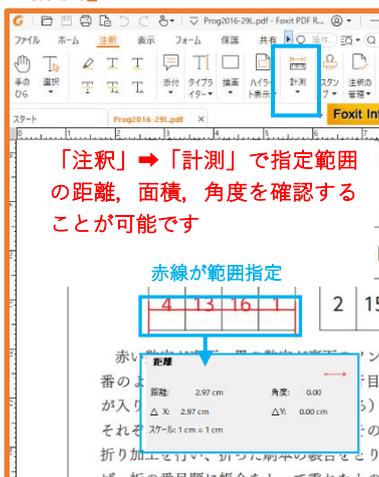


Foxtit = 「表示」⇒「支援」⇒「ルーペツール」

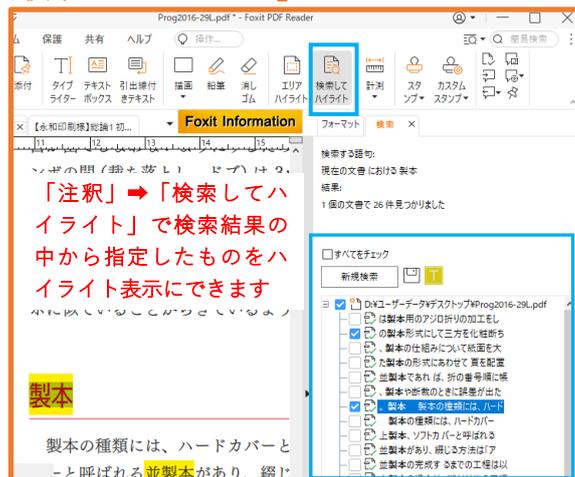
Acrobat = 「表示」⇒「ズーム」⇒「ルーペツール」

比較 6 その他 Foxtit PDF Reader のみにある機能 (比較 5 の拡大表示機能をご確認ください)

「計測」



「検索してハイライト」



「文字カウント」

