

ソフト名称	類字抽出ツール
実行ファイル	RuijiFilter.exe
本書の版数	初版 2019年 8月22日作成
著作権	H R C

1. 機能紹介 (主な機能)

このソフトは、類字の抽出に特化した Windows用のツールです。
類字とは、同じ意味の言葉で全角／半角、ひらがな／カタカナ、大文字／小文字による表示が異なる文字列のことで、以下の事例があります。

NO	データ名称	備考
1	シャーベット	全て全角のカタカナ("シャ" と "ベ" の間は全角の長音符)
2	シャーベット	NO.1 の "シャ" と "ベ" の間が半角のハイフン
3	シャ-ベッ	NO.1 の 全ての文字が半角のカタカナ
4	シャーベッ	NO.1 の "ベ" の濁点が "ヘ" と分離して 2文字に分裂
5	しゃべつと	NO.1 の 全ての文字が全角のひらがな
6	しゃべつと	NO.5 の "ゃ" が 全角 & 大文字のカタカナ

例えば、表計算ソフトで上記の事例を含んだデータベースを作成している場合、人間は上記を全て同じ意味の文字として認識できますが、集計等の作業をシステム処理する場合は問題(※)となる可能性があります。

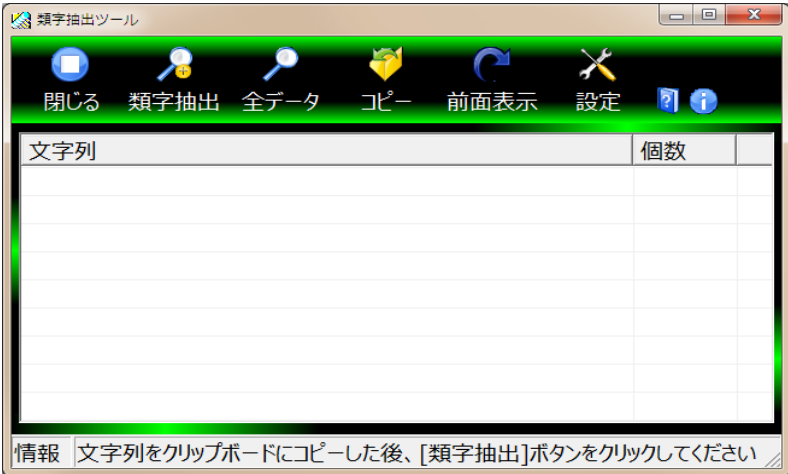
※問題例

同じ意味のデータが、別個のデータとしてシステム処理される
上記事例の場合、表計算ソフトのカウント関数等で "シャーベット"の個数を数えると
意味的な結果=6 に対して、カウント結果=1 となる

この問題は、複数の類字を一つの文言に統一すれば解決します。
そのためには、まず類字の有無を確認することが必要です。
類字の有無は、表計算ソフトのフィルタ機能でも確認できますが、データが大量の場合、その確認作業は大きな手間がかかります。

本ソフトは、大量のデータから類字のみを抽出して確認する作業を、ごく簡単な操作で自動処理するツールとしてご利用いただけます。

(本ソフトの外観)



1. 機能紹介 (活用事例)

下の表データの緑のセルをコピーして、本ツールを実行した場合

NO	データ名称	備考
1	アップルパイ	全て全角のカタカナ
2	エクレア	同上
3	クッキー	同上
4	しゃーべつと	全て全角のひらがな
5	しゃーべつと	NO.4 "ゃ" が 全角 & 大文字のカタカナ
6	シャ-ベツト	全角のカタカナ & 長音符"ー" が 半角のハイフン "-"
7	シャ-ヘ`ツト	"ベ" の濁点が "ハ" と分離して 2文字に分裂
8	シャ-ベツト	全て全角のカタカナ
9	シャ-ベツト	NO.8 と同一
10	シャ-ハ`ツト	NO.8 の全ての文字が半角カタカナ

実行例-1/2 ※類字のみを抽出して表示

類字(NO.4～10)のみを抽出して表示します。
NO.1～3のデータは類字が存在しないため、表示されません。
このため、類字のみに焦点を絞った確認が簡単にできます。
表計算ソフトでデータベース等を作成していて、文字列の入力規則を厳密に設定することが難しい場合、この類字抽出機能を活用することで、不適切な類字の混在を簡単に確認できます。
※類字の統一は、表計算ソフトの編集機能で不適切な類字 → 正しい文字列に一括置換すればよいでしょう

実行例-2/2 ※全てのデータを確認

全てのデータの種類と個数を、ワークシート関数等を使用せず、すぐに確認できます。
また、データ文字列の重複が許されない場合(例. シリアル番号等)、重複しているデータをすぐに発見することができます。重複データを発見する手順は、[4-3. 操作方法 (ソート：データの並べ替え)]をご覧ください。

2. 動作環境

Windows-7, 10(32,64ビット)上で動作します。

※ Windows-8でも動作すると思われますが、動作確認はしていません

3. インストールと起動

以下×2ステップの手順でインストール → 起動してください。

ステップ-1：本ソフトの圧縮ファイルをどこか適当なフォルダに解凍します。

ステップ-2：実行ファイル(RuijiFilter.exe)を起動してください。

なお、アンインストールされる場合は、[11. アンインストールの手順] をご覧ください。

4-1. 操作方法 (類字抽出)

表計算ソフトの文字列データをクリップボードにコピー(※)した後、[類字抽出]ボタンをクリックすると、類字のみを表示します。

類字のないデータは表示されないの、類字のみに焦点を絞った確認が容易にできます。



※表計算ソフトの文字列データを簡単にコピーする手順を[4-6. 操作方法(表計算ソフトのデータを簡単にコピーするヒント)]で紹介していますので、あわせてご覧ください

なお、データに類字が存在しない場合は、何も表示されません。

不適切な類字が混在している場合は、表計算ソフト等でデータを一括置換等で修正した後、本ツールの類字抽出で類字が表示されなくなることを確認してください。

4-2. 操作方法 (全データの表示)

[全データ]ボタンをクリックすると、全てのデータの文字列内容と個数が表示されます。



重複データを発見する手順は、[4-3. 操作方法 (ソート：データの並べ替え)]をご覧ください。

4-3. 操作方法 (ソート：データの並べ替え)

結果リストのカラム(データ項目のタイトルを表示している部分)をクリックすると、データをソートして再表示します。

[文字列]カラムをクリックした場合

The screenshot shows the '類字抽出ツール' (Similar Word Extraction Tool) interface. On the left, the 'Text' column header is highlighted with a red box and a red arrow. A blue arrow points to the right, showing the result after clicking. The data is now sorted alphabetically by the text values.

文字列	個数
アッブルパイ	1
エクレア	1
クッキー	1
しゃべっと	1
しゃべっと	1
しゃべっと	1
しゃべっと	1
しゃべっと	1
しゃべっと	1
しゃべっと	2
しゃべっと	1

情報 全10個のデータの中から、9種類のデータを取得しました

[個数]カラムをクリックした場合

The screenshot shows the '類字抽出ツール' (Similar Word Extraction Tool) interface. On the left, the 'Count' column header is highlighted with a red box and a red arrow. A blue arrow points to the right, showing the result after clicking. The data is now sorted by the count values in descending order.

文字列	個数
しゃべっと	2
アッブルパイ	1
エクレア	1
クッキー	1
しゃべっと	1
しゃべっと	1
しゃべっと	1
しゃべっと	1
しゃべっと	1
しゃべっと	1
しゃべっと	1

情報 全10個のデータの中から、9種類のデータを取得しました

重複データを発見する場合は、個数でデータをソートして、個数が1より大きいデータをリストの最上位に表示させれば、簡単に確認できます。

4-4. 操作方法 (表示しているデータをコピー)

表示している全てのデータをコピーする場合は、[コピー]ボタンをクリックしてください。

コピーしたデータは、表計算ソフト等に貼付けることができます。

※同様の機能は、リストの右クリックで登場するポップアップメニューでも処理できます

The screenshot shows the '類字抽出ツール' (Similar Word Extraction Tool) interface. The 'コピー' (Copy) button in the toolbar is highlighted with a red box and a red arrow.

文字列	個数
アッブルパイ	1
エクレア	1
クッキー	1
しゃべっと	1
しゃべっと	1
しゃべっと	1
しゃべっと	1
しゃべっと	1
しゃべっと	1
しゃべっと	2
しゃべっと	1

情報 全10個のデータの中から、9種類のデータを取得しました

(ポップアップメニュー 外観)

- 全てのデータをクリップボードにコピーする
- 選択したデータをクリップボードにコピーする
- 選択したデータをリストから削除する

[選択したデータをリストから削除する]メニューの詳細は[4-6. 操作方法 (表計算ソフトのデータを簡単にコピーするヒント)]をご覧ください。

コピーの対象は、①文字列のみ ②文字列 + 個数の いずれかを設定できます。

設定の詳細は [5-4-3. 表示している文字列をコピーする際、[個数]もコピーする]をご覧ください。

4-5. 操作方法 (前面表示)

表計算ソフトと本ツールを交互に操作する際、本ツールを常に最前面に表示させておくと、操作が楽になります。

上記の操作をされたい場合は、[前面表示]ボタンをクリックしてください。



※本ツールを最前面に表示させた場合
表計算ソフトを操作しても、本ツールは常に表示された状態になります

	A	B	C	D
1				
2		■ データリスト		
3		NO	データ名称	
4		1	アップルパイ	全て全角の
5		2	エクレア	同上
6		3	クッキー	同上
7		4	しゃーべっと	全て全角の
8		5	しゃーべっと	NO.4 "や"
9		6	しゃーべつ	全角のカタカナ & 長音符"ー" が 半角のハイフン "-"
10		7	しゃーへゝつ	"べ" の濁点が "へ" と分離して 2文字に分裂
11		8	しゃーべつ	全て全角のカタカナ
12		9	しゃーべつ	NO.8 と同一
13		10	しゃーへゝつ	NO.8 の全ての文字が半角カタカナ



最前面表示を解除する場合は、[通常表示]ボタンをクリックしてください。



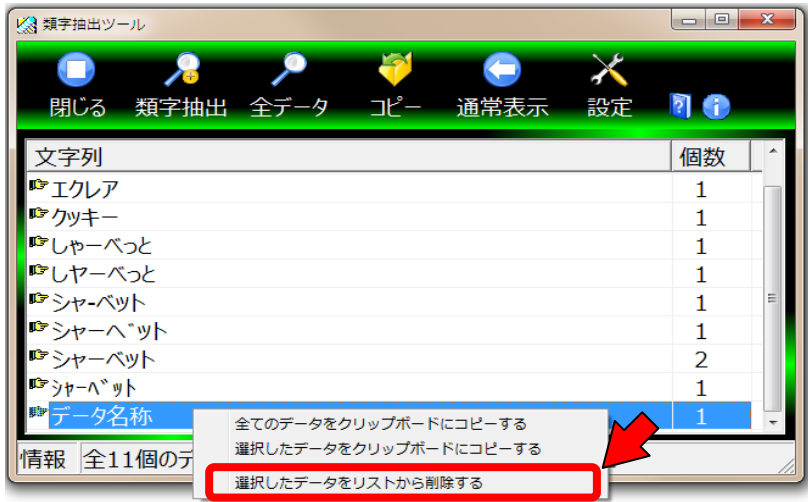
4-6. 操作方法 (表計算ソフトのデータを簡単にコピーするヒント)

表計算ソフトのデータをコピーする際、データのセル部分のみを選択する作業は意外と面倒です。
そこで表計算ソフトのカラムをクリックすると、全ての行のデータを一括で選択することができますので、データを簡単にコピーすることができます。



	A	B	C	D
1				
2		■ データリスト		
3		NO	データ名称	備考
4		1	アップルパイ	全て全角のカタカナ
5		2	エクレア	同上
6		3	クッキー	同上
7		4	しゃーべっと	全て全角のひらがな
8		5	しゃーべっと	NO.4 "ゃ" が 全角 & 大文字のカタカナ
9		6	シャ-ベツト	全角のカタカナ & 長音符"ー" が 半角のハイフン "-"
10		7	シャ-ヘ`ツト	"ベ" の濁点が "ヘ" と分離して 2文字に分裂
11		8	シャ-ベツト	全て全角のカタカナ
12		9	シャ-ベツト	NO.8 と同一
13		10	シャ-ヘ`ツト	NO.8 の全ての文字が半角カタカナ

この場合、表タイトルの文字列(上記事例の場合は "データ名称")がデータに含まれてしまいますが、タイトル文字列を無視するか、あるいは本ツールでデータを取得した後、表示を削除することで、その後の確認作業に影響はなくなります。
本ツールがリスト表示しているデータを削除するには、削除するデータを選択した後、右クリックで登場するポップアップメニューの[選択したデータをリストから削除する]メニューを実行するか、[Delete]キーを押してください。



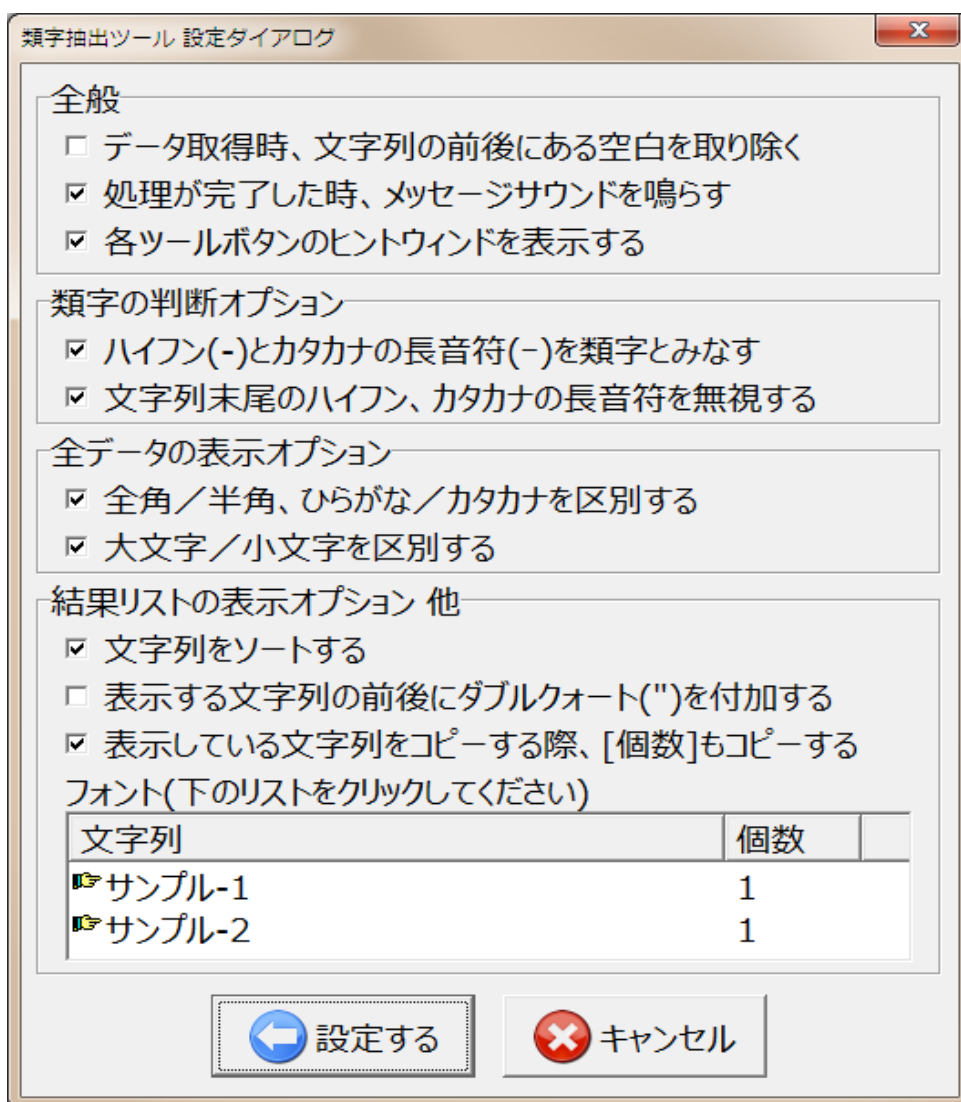
5. 設定 (設定ダイアログの呼び出し)

本ツールの動作環境を設定するには、[設定]ボタンをクリックしてください。



これで、下図の設定ダイアログウィンドが登場するので、後述する各種設定機能を確認されて、お好みにあわせた動作環境を設定してください。

※下図はデフォルト(初期設定)の状態です



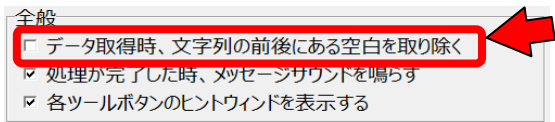
設定内容を変更した後、[設定する]ボタンをクリックすると内容が保存されます。
いったん変更した内容を破棄する場合は、[キャンセル]ボタンをクリックしてください。

5-1. 設定 (全般)

以下×3項目を設定できます。

5-1-1. データ取得時、文字列の前後にある空白を取り除く

取得した文字列の前後に空白がある場合、その空白を取り除く設定で、デフォルトはオフ(チェックマークがなくて無効)です。



この設定と下図のサンプルデータの場合の動作結果は以下になります。

NO	データ名称	備考
1	X123-4567	全て半角の英数字
2	X123-4567	NO.1 の文字列の先頭に半角の空白がある

※動作結果

設定	類字抽出した時	全データを表示した時
オフ		
オン		

※補足説明

この設定がオフの場合、NO.2の文字列先頭の空白は取り除かれずにデータ処理されるため、NO.1 と NO.2は異なる文字列として扱われます。

※ただし、それぞれは類字として抽出されます

一方オンの場合、NO.2の先頭空白は取り除かれてデータ処理されるため、NO.1と NO.2は同一文字列として扱われます。

このため、それぞれは類字として認識されず、また全データを表示した場合も同一文字列として個数がカウントされます。

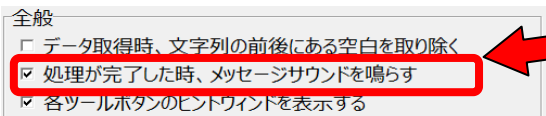
よって、この設定は文字列前後の空白をチェックする場合はオフ、空白に留意する必要がない場合はオンにしてください。

なお、この設定をオフにして、もし仮に文字列の末尾に空白があった場合、表示をただけでは空白を識別できません。この場合は、[5-4-2. 表示する文字列の前後にダブルクォート(")を付加する]の設定をオンにすることをお勧めします。

5-1. 設定 (全般)

5-1-2. メッセージサウンド

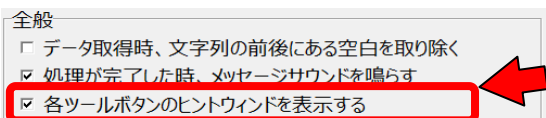
本ツールが処理(類字の抽出処理等)を完了したことを音で知らせる設定です。



デフォルトはオンですが、音が不要の場合はオフにしてください。

5-1-3. ヒントウィンド

各ツールボタンの上でマウスカーソルを停止すると、そのボタンの機能を説明するヒントウィンドを表示する設定です。

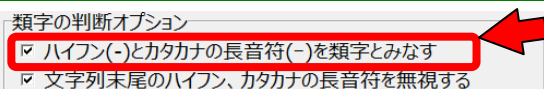


デフォルトはオンですが、ヒント表示が不要の場合はオフにしてください。

5-2. 設定 (類字の判断オプション)

以下×2項目を設定できます。

5-2-1. ハイフン(-)とカタカナの長音符(ー)を類字と見なす



たとえば、①"データ" ②"デ-タ" という文字列があったとします。

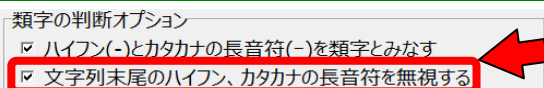
※それぞれの違いは、"デ"と"タ"の区切りが ①長音符、②ハイフンで異なっている点です

この設定をオンにして類字抽出した場合、①②は類字としてそれぞれ表示されます。一方、オフの場合、①②はそれぞれ別個の文字列として処理されるため、類字として表示されなくなります。

なお、全データを表示した場合は、この設定のオン／オフに関わらず、①と②は別個のデータとして表示されます。

デフォルトはオンですが、ニーズに応じて適宜設定をオフにしてください。

5.2-2. 文字列末尾のハイフン、長音符を無視する



たとえば、①"プリンタ" ②"プリンター" という文字列があったとします。

この設定をオンにして類字抽出した場合、①②は類字としてそれぞれ表示されます。一方、オフの場合、①②はそれぞれ別個の文字列として処理されるため、類字として表示されなくなります。

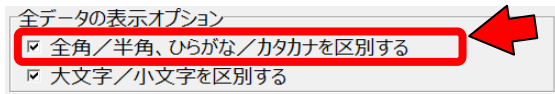
なお、全データを表示した場合は、この設定のオン／オフに関わらず、①と②は別個のデータとして表示されます。

デフォルトはオンですが、ニーズに応じて適宜設定をオフにしてください。

5-3. 設定 (全データの表示オプション)

5-3-1. 全角／半角、ひらがな／カタカナを区別する

取得した全てのデータを表示する際、全角と半角、ひらがなとカタカナを区別して個数をカウントする設定で、デフォルトはオンです。



この設定と下図のサンプルデータの場合の動作結果は以下になります。

NO	データ名称	備考
1	X123-4567	全て半角の英数字
2	X 1 2 3 - 4 5 6 7	NO.1 の全ての文字が全角の英数字
3	X123-4567-A	全て半角の英数字
4	あいうえお	全て全角のひらがな
5	アイウエオ	NO.4 の全ての文字が全角のカタカナ
6	アイウイ	// 半角の //

※動作結果

設定	類字抽出した時	全データを表示した時
オン		
オフ	上と同じ	

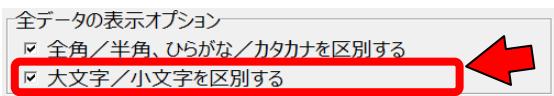
※補足説明

この設定は、全データの表示のみに適用されます。
設定がオンの場合、NO.1,2と4,5,6の文字列はそれぞれ区別されて表示されますが
オフの場合、上記の組み合わせは区別されない = 同一文字列と判断されます。
その場合、代表的な文字列ならびに同一文字列と判断された個数の合計が
表示されます。

5-3. 設定 (全データの表示オプション)

5-3-2. 大文字／小文字を区別する

取得したデータの全データを表示する際、大文字と小文字を区別して個数をカウントする設定で、デフォルトはオンです。



この設定と下図のサンプルデータの場合の動作結果は以下になります。

NO	データ名称	備考
1	X123-4567	全て半角の英数字
2	x123-4567	NO.1 の最初の一文字が半角小文字のエクス(x)
3	X123-4567-A	全て半角の英数字

※動作結果

設定	類字抽出した時	全データを表示した時
オン		
オフ	上と同じ	

※補足説明

この設定は、全データの表示のみに適用されます。
設定がオンの場合、NO.1,2の文字列はそれぞれ区別されて表示されますが
オフの場合、上記の組み合わせは区別されない = 同一文字列と判断されます。
その場合、代表的な文字列ならびに同一文字列と判断された個数の合計が
表示されます。

5-4. 設定 (結果リストの表示オプション 他)

5-4-1. 文字列をソートする

データを表示する際、文字列をソートして表示する設定で、デフォルトはオンです。

結果リストの表示オプション 他

☒ 文字列をソートする

☐ 表示する文字列の前後にダブルクォート(")を付加する

☒ 表示している文字列をコピーする際、[個数]もコピーする

フォント(下のリストをクリックしてください)

文字列	個数
📁 サンプル-1	1
📁 サンプル-2	1

この設定と下図のサンプルデータの場合の動作結果は以下になります。

NO	データ名称	備考
1	X123-4567-C	全て半角の英数字
2	X123-4567-B	//
3	X123-4567-A	//

※動作結果(全データ表示)

設定：オン

オフ

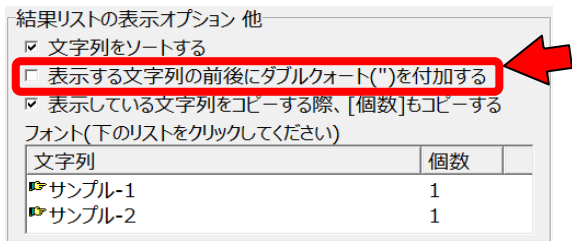
※補足説明

コピー元(表計算ソフトのオリジナルデータ等)と同じ順序でデータを確認する場合はこの設定をオフにしてください。

5-4. 設定 (結果リストの表示オプション 他)

5-4-2. 表示する文字列の前後にダブルクォート(")を付加する

表示する文字列の前後にダブルクォート(")を付加する設定で、デフォルトはオフです。



この設定と下図のサンプルデータの場合の動作結果は以下になります。

NO	データ名称	備考
1	X123-4567	全て半角の英数字
2	X123-4567	NO.1 の文字列の末尾に半角の空白がある

※動作結果(全データ表示)

設定：オフ	オン

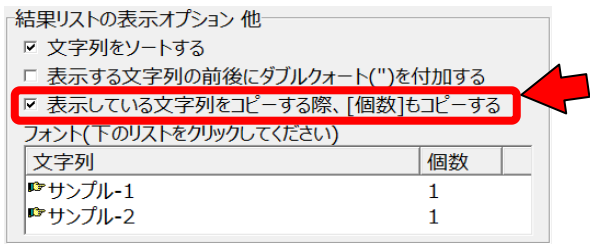
※補足説明

この設定をオンにすると、[文字列の前後空白を取り除く]の設定をオフにして、文字列の前後に空白が含まれているケースがあった場合、それを一目で確認できます。

5-4. 設定（結果リストの表示オプション 他）

5-4-3. 表示している文字列をコピーする際、[個数]もコピーする

[コピー]ボタンをクリックして、表示している文字列をコピーする際、個数もコピーする設定で、デフォルトはオフです。



下図の表示がされている場合の動作結果は以下になります。



※動作結果(コピーされるデータ)

設定：オン		オフ
X123-4567-A	1	X123-4567-A
X123-4567-B	1	X123-4567-B
X123-4567-C	1	X123-4567-C

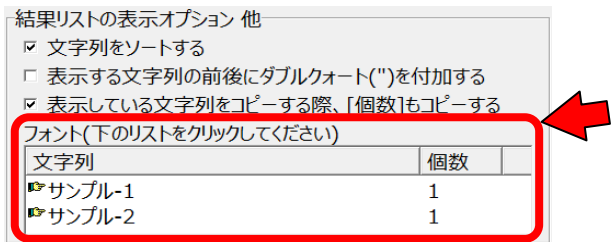
※補足説明

この設定をオンにして、[文字列] と [個数]をコピーした場合、それぞれのデータの間にタブ記号を挿入しますので、表計算ソフト等に結果を貼付けると、それぞれのデータは別のセルに保存されます。

5-4-4. 結果リストの表示フォント

結果を表示しているリストのフォントを設定できます。

サンプル表示されているリストをクリックすると、フォントを選択するダイアログが登場するので、好みのフォントをお選びください。



6. 仕様制約

ある行の最初の文字がダブルクォート(")で始まって、かつ、その次の行の末尾がダブルクォートになっている場合は、1行の文字列と判断します。

※事例

1行目	"AB
2行目	C"

この事例の場合、"ABC"という文字列が1行に収容されていると判断します。

これは、表計算ソフトの改行仕様に合わせているためです。

また、特殊文字(●と○)や漢数字(壺と1)の類字判断には対応していません。

7. 作業ニーズに応じた本ソフトの個別カスタマイズについて

本ソフトは、操作の扱いやすさと、一般的なデータ処理ニーズへの対応を最優先して開発しています。

従って、ご利用される方々独自の細かい作業ニーズから見ると、機能が不足している点が多々あると思われます。

上記不足ニーズに対する機能追加のご要望があれば、個別に開発対応いたします。その場合、有償とさせていただきますが、格安で対応いたしますので、まずはお気軽にメールで作者にご相談ください。

作者のメールアドレスは[8. 作者への問い合わせ先]をご覧ください。

(個別カスタマイズの事例)

■ケース1：個別ルールに従った文字列のチェック

文字列中の定型記号("/"等 - 位置不定)の後の文字列が指定の文字列になっている or なっていないデータのみを抽出

(事例: "/" の後の文字列が "JP" になっているものだけを抽出する等)

■ケース2：個別ルールに応じた文字列の置換処理

ケース1のルールに合致した文字列を、指定したルールに従った文字列に自動置換

■ケース3：暗号化

指定したセルの文字列を暗号化

(データファイル全体のうち、任意の一部を特定者にのみ開示する用途向け)

※上記の事例以外についても、柔軟に対応させていただきます

日々大量の文字列データを扱われている場合、あなた専用の個別カスタマイズ版をご利用いただくことで、作業効率を大きく飛躍させていただければ幸いです。

8. 作者への問い合わせ先

個別カスタマイズの開発対応は有償とさせていただきますが、一般公開バージョンの不具合の改修、汎用的な機能改善(多くのユーザの皆様に通じてご利用いただける機能の追加等)については、無償で対応いたします。

ご要望があれば、作者にご遠慮なくメールでお問い合わせください。

作者メールアドレス: peb00732@nifty.com

※上記メアドの "00" は数字のゼロです

9. ご利用条件

[10. 免責事項]にご同意いただける場合に限り、本ソフトをご利用してください。
 なお、ご同意いただけない場合、本ソフトのご利用はおやめいただき、本ソフトをご使用のPCからアンインストールしていただく様、お願いします。
 ※アンインストールの手順は[11. アンインストールの手順] をご覧ください。

10. 免責事項

本ソフトの動作不具合等により、ご利用される方にいかなる損害が発生したとしても作者は一切の責任を負いません。
 また、ソフトの汎用的な機能改善に関するご要望をいただいた場合、作者はできるだけご要望に沿う対応努力はするものの、義務的な責務は負いません。

11. アンインストールの手順

以下×四つのファイルをエクスプローラ等で削除すればアンインストールは完了です。

- ① RuijiFilter.exe
- ② RuijiFilter.ini
- ③ RuijiFilter.pdf
- ④ Readme(RF).txt

※本ソフトはご使用PCのシステム領域(レジストリ等)に一切アクセスしていません

12. 著作権

このソフトの著作権は、作者の H R C が有しています。
 Copyright (c) 2019 H R C AllRights Reserved.

13. 転載条件

ご自由に転載(配布)していただいてもかまいません。
 なお、転載された場合は、事後でよいので、その旨を作者にメールでお知らせください。

14. 謝辞

下記のプログラムリソースを利用させていただきました。
 ご提供いただいた作者の皆様に深くお礼申し上げます。

プログラムリソース	作者 様
TFlipImage	河邦 正 様
ツールボタン用画像	Aha-Soft 様 ※ http://www.aha-soft.com/jp/index.htm