

# 「コンピュータの舞台裏」

## 第2回

### 「業務への正規表現の応用」

2015-12-19

By くーへん

<https://github.com/flhtc1964/Computer>

# (自己 | 事故) 紹介

- 1964年 東京都生まれ
- N88BASIC、MSDOS、VBA、VBS、PHP、C#
- SunOS4.1.3、FreeBSD  
Cシェル、grep、gawk、sed、Perl
- UUCP、NetNews、kermit
- 2000年 4月から大阪の某法律事務所へ転職  
(離婚したし、転職のお誘いがあったので)  
情報システム課に所属し会社更生システム(SQL  
Server + Access[adp])の作成、データベース作成  
&メンテナンスと、主にOffice製品等のヘルプデ  
スクを担当現在に至る

# Unix(ユニックス) OS

- Windows OSが世に出る前は  
Unix(ユニックス)という  
オペレーティングシステム(OS)が動く  
サーバがインターネット上で複数台稼働  
しており利用者はそのサーバにTelnetコマンドでサー  
バにログインし、sh(シェル)という  
対話型のコマンドを使ってメールやNetNews  
そしてテキストデータ処理等を行っていた  
<https://ja.wikipedia.org/wiki/UNIX>
- Unixのsh(シェル)とは、WindowsのMS-DOSの  
「コマンドプロンプト」だと思ってください

# 正規表現

## (せいきひょうげん)

- 正規表現は、**Unixのお陰で一般化**された
- 大量のテキストデータに対し統一した検索ルール(**パターンマッチ**)が必要となり

一括検索:	grep(egrep)
一括(検索 置換)、一括削除:	sed
一括レコード単位処理:	awk(gawk)

C言語、sed、awk、sh(シェル)の  
機能を取り入れたプログラム言語: Perl

これらは「正規表現」を使って処理をしている  
最近ではWindowsで動くエディタの検索or置換時に「正規表現」  
の利用が可能



# 正規表現メモ

<http://www.kt.rim.or.jp/~kbbk/regex/regex.html>

## • 2012年8月現在の 「正規表現」使用情報

### 正規表現メモ

最終更新日 2012年8月24日

正規表現は使い慣れれば便利なものですが、ツールによって使える正規表現演算子(メタキャラクタ)に違いがあったりして戸惑うこともあります。そこで、正規表現を扱うツールの代表的なものをいくつか選び、そこで使われている正規表現演算子をまとめてみました。

### 正規表現入門者/初心者の方へ

正規表現のチュートリアルが、[perlの正規表現チュートリアル](#)にあります。Perl5.8のドキュメントですが、一般的な入門にも使えると思います。

### Table of contents

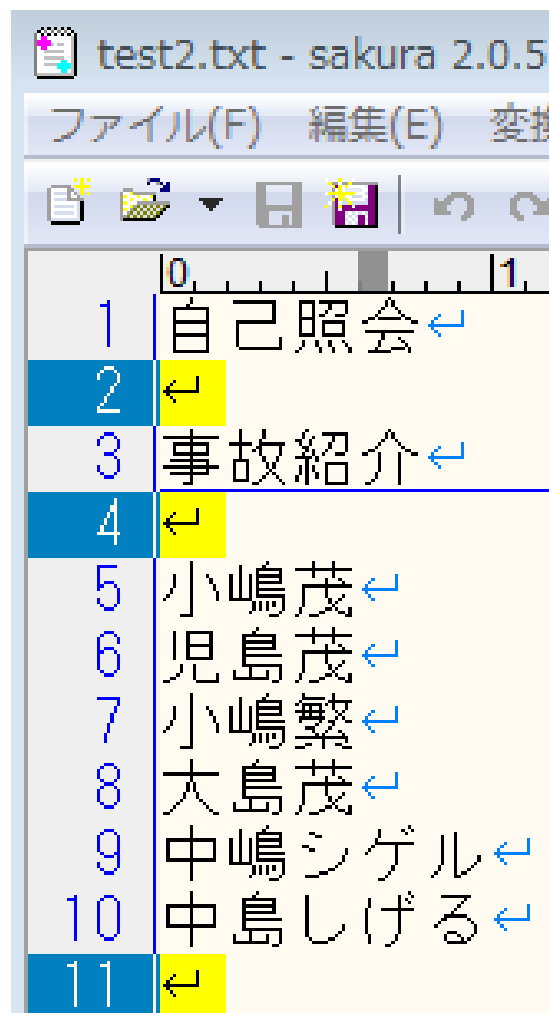
1. [grep](#)で使える正規表現
2. [egrep](#)で使える正規表現
3. [sed](#)で使える正規表現
4. [awk](#)で使える正規表現
5. [Perl](#)で使える正規表現
6. [Python](#)で使える正規表現
7. [Ruby](#)で使える正規表現
8. [gawk 3.0\(以降\)](#)で使える正規表現
9. [Tcl 8.2.3\(以降\)](#)で使える正規表現
10. [PCRE](#)で使える正規表現
11. [PHP](#)で使える正規表現([mb\\_ereg](#))
12. [.NET Framework](#)で使える正規表現
13. [Java\(1.4以降\)](#)で使える正規表現
14. [POSIX 1003.2](#)での正規表現について
15. [各正規表現演算子の説明](#)
16. [各エスケープシーケンスの説明](#)
17. [各処理系正規演算子一覧表](#)
18. [ある文字列を含まない正規表現](#)
19. [リンク](#)
20. [用語集](#)
21. [参考文献](#)

# テキスト一括置換「メタ文字」 「空白行」を指定して→削除

- 行頭 → ^
- 行末 → \$
- 行頭 + 行末 = 空白 = ^\$
- 改行 = ¥r¥n
- ^\$ ¥r¥n = 空白改行

注意: ¥r¥nは使用する  
正規表現エンジンにより  
指定しない事もあります

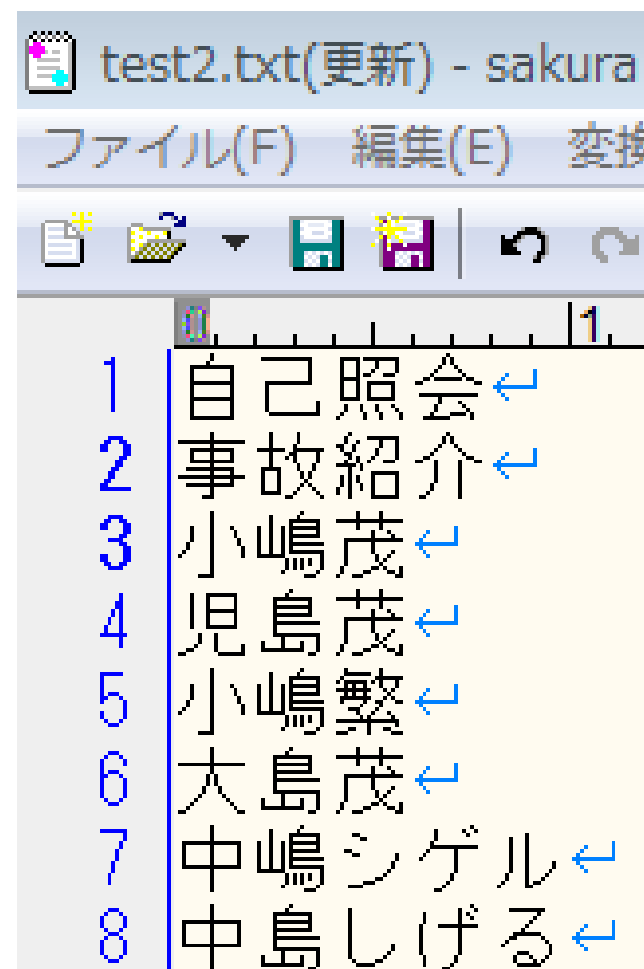
置換	
置換前(N)	^\$¥r¥n
置換後(P)	
<input type="checkbox"/>	クリップボードから貼り付ける(T)
<input type="checkbox"/>	単語単位で探す(W)
<input type="checkbox"/>	英大文字と小文字を区別する(O)
<input checked="" type="checkbox"/>	正規表現(E)



置換

置換前(N) ~\$¥r¥n

置換後(P)



# 誤記のパターンを指定し 「どちらか」だったら正しく置換

- 誤パターン → 自己照会
- 誤パターン → 事故紹介
- 正しい表記 → 自己紹介

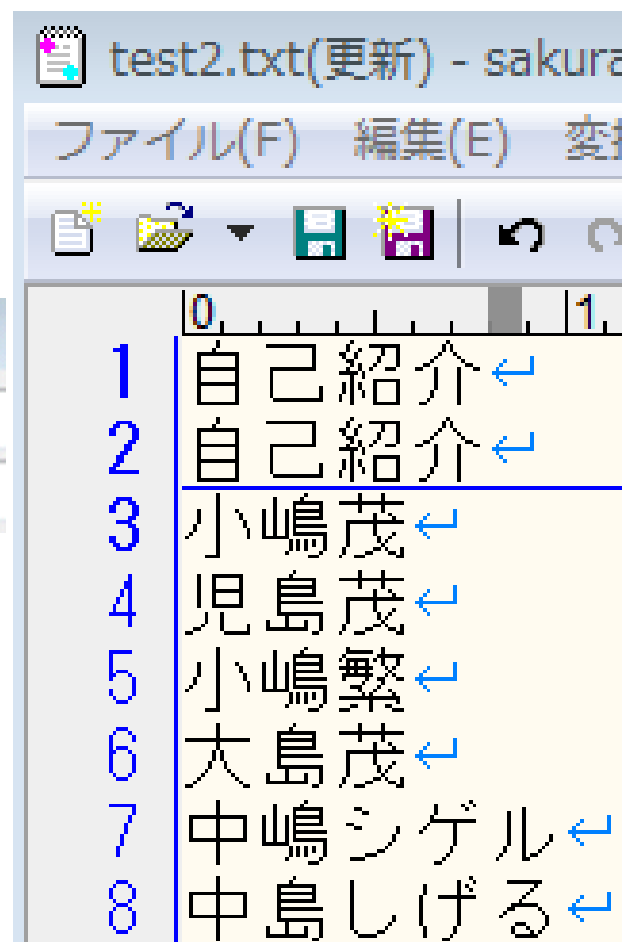
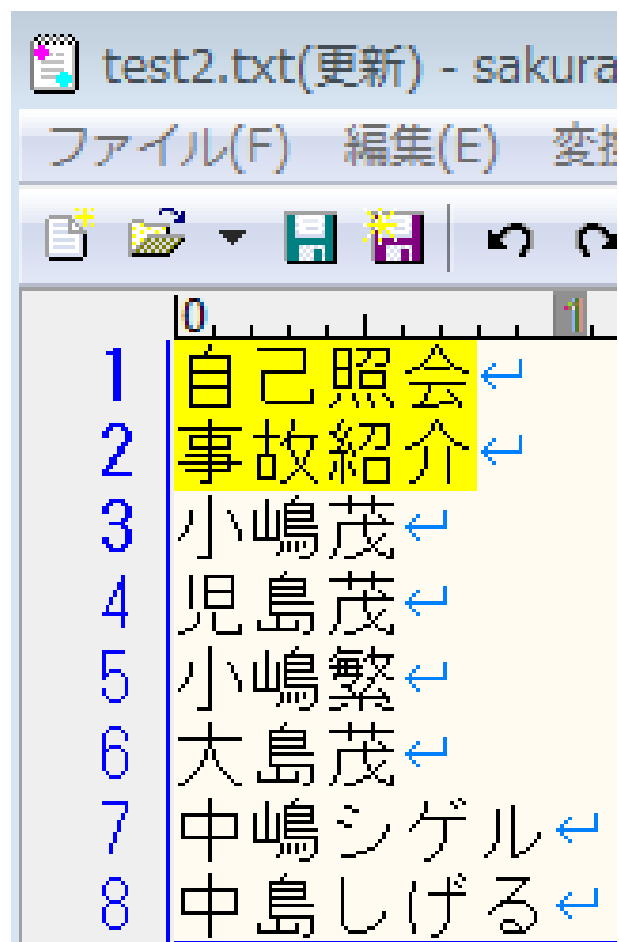
## 置換

置換前(N) 自己照会|事故紹介

置換後(P) 自己紹介

- ☐ クリップボードから貼り付ける(T)
- ☐ 単語単位で探す(W)
- ☐ 英大文字と小文字を区別する(C)
- ☒ 正規表現(E)





# 例：3文字の組み合わせにマッチ

test2.txt(更新) - sakura 2.0.5.0

ファイル(F) 編集(E) 変換(C) 検索(S) ツール(T) 設定(O) ウ

置換

置換前(N) [小児大][島嶋][茂繁]

置換後(P) 小島茂

☐ クリップボードから貼り付ける(T)

☐ 単語単位で探す(W)

☐ 英大文字と小文字を区別する(C)

☒ 正規表現(E)

bregonig.dll Ver 3.02 with Onigmo 5.13.2

1 自己紹介←

2 自己紹介←

3 小嶋茂←

4 児島茂←

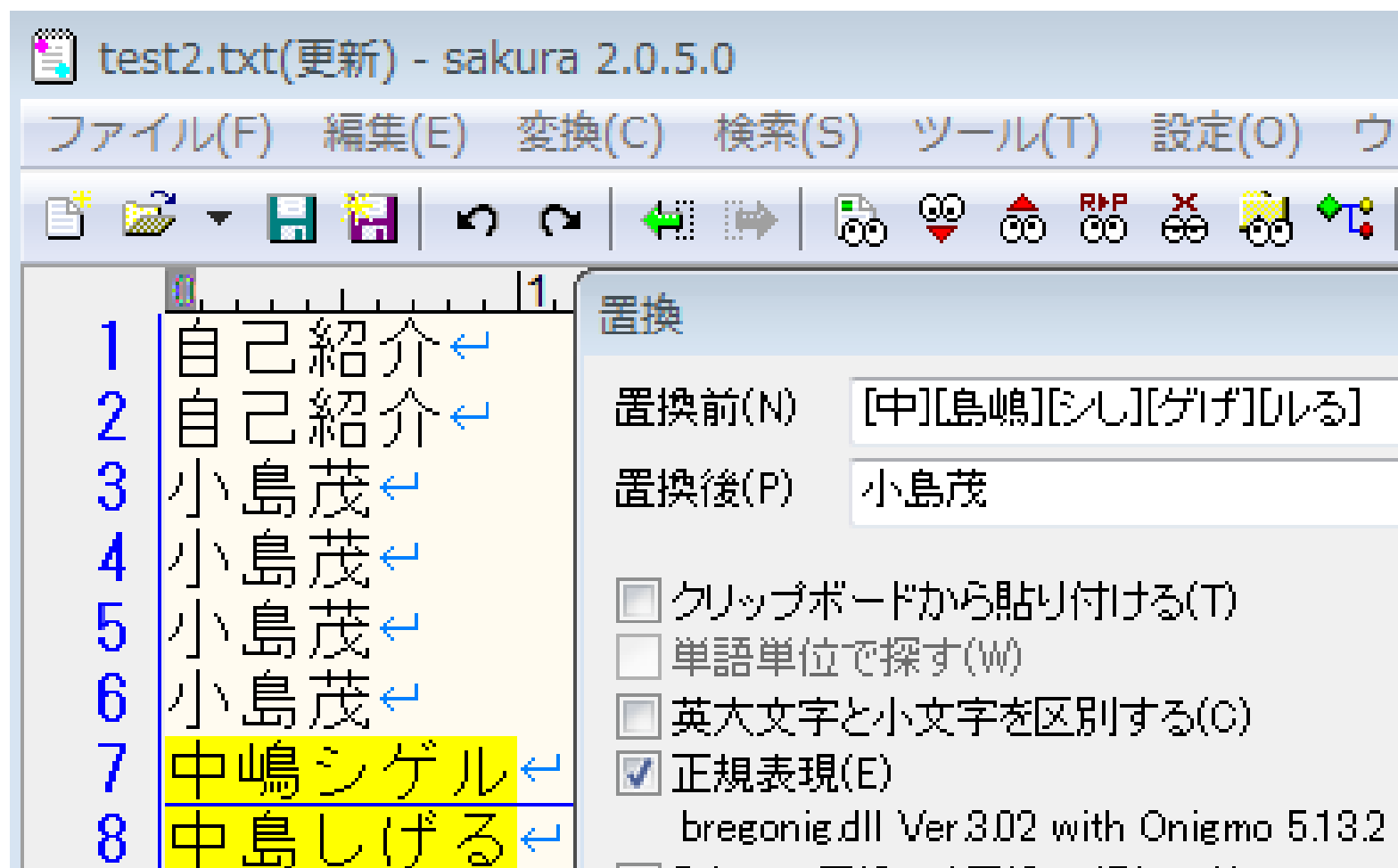
5 小嶋繁←

6 大島茂←

7 中嶋シゲル←

8 中島しげる←

# 例: 5文字の組み合わせにマッチ



# 失敗しない為には

- 複数の抽出条件を指定して絞り込みたい時は → 1つの条件で絞り込んだ結果をまずファイルへ出力(保存)する
- 保存したファイルを → 条件で絞り込む
- 正規表現には「方言」があるのでネット等で調べて使用すること
- ( や [ 等のメタ文字は ~~¥~~( や ~~¥~~[  
の様に「¥」を付加する場合がある

# Windowsで動く onigsed.exe(20091031版)

GNU sed 4.1.5 の日本語 Windowsへの移植(というほどたいしたことはしていない)です。

cygwinに付属のGNU sedがMBCS対応していないので作業しました。

VC++ 7.1でコンパイルしています。特にDLLは必要ありません。

また、正規表現エンジンとして鬼車(2.5.0)を組み込んでおり、オプション指定によりPerl互換の正規表現を使うことができます。で

きるだけオリジナルと同じ振る舞いをするように努力しましたが、どこかに抜けがあるかもしれません(特に改行の扱いに関して)。

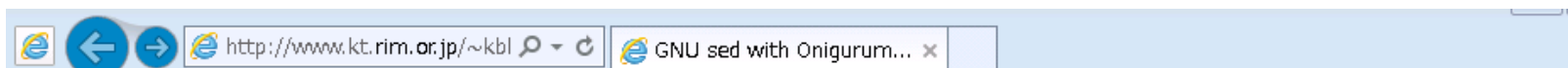
漢字コードとしてshiftjis、euc-jp、utf-8 が使えます。使用するときにはコマンドラインオプションで

- shiftjis -Wctype=SJIS または --ctype=SJIS (デフォルト)
- euc-jp -Wctype=EUC または --ctype=EUC
- utf-8 -Wctype=UTF8 または --ctype=UTF8

マルチバイト処理をオフにしたい場合には -Wctype=ASCII または --ctype=ASCII を指定します。

[readme ja](#) より抜粋

# http://www.kt.rim.or.jp/~kbbk/sed/



## GNU sed with Oniguruma (Onigsed)とは?

正規表現エンジンとして、[鬼車](#)を組み込んだ sed です。オプション指定により、POSIX EREや Perl5互換の正規表現を使うことができます。

### ダウンロードリンク

- [Onigsed 2009年10月31日バージョン](#)  
md5 aae0a510dee4de6acca29cc9c86ca4a5 \*onigsed-20091031.zip  
sha1 2738e55c05883c334c66dcb264b5ab59202bf8ca onigsed-20091031.zip
- [Onigsed 2009年1月1日バージョン](#)  
9728a92c58ef001984a9fed51095535f \*onigsed-20090101.zip  
0b93c97ba2781a44800e1d888239ac67ec54c62a5f3485196be7d757ebaaf873 \*onigsed-20090101.zip
- [鬼車を使っていないバージョン\(2009/10/5\)](#)  
md5 3fd753202cb5dbc565eeba29abd049d \*sed-mbcs-win32-20091004.zip  
sha1 9f4f44f67cb1fa6ad9c95fa3ac232be625b36f6 sed-mbcs-win32-20091004.zip  
[\(2009/5/25\)版](#)  
949ec827c9383d9ce60c947115db0cdf \*sed-mbcs-win32-20090525.zip  
8bdc396588ff1e978562e53fbff22e63d5703903 sed-mbcs-win32-20090525.zip

以下、readme.jaより。

・これはなに？

GNU sed 4.1.5 の日本語 Windowsへの移植(というほどたいしたことはしていない)です。cygwinに付属のGNU sedがMBCS対応していないので作業しました。VC++ 7.1でコンパイルしています。特にDLLは必要ありません。また、正規表現エンジンとして鬼車(2.5.0)を組み込んでおり、オプション指定によりPerl5互換の正規表現を使うことができます。できるだけオリジナルと同じ振る舞いをするように努力しましたが、どこかに抜けがあるかもしれません(特に改行の扱いに関して)。

# onigsed.exeで置換してみる

The image shows a Windows environment with two Notepad windows and a Command Prompt window.

**test2.txt - メモ帳**

自己照会  
事故紹介  
小嶋茂  
児島茂  
小嶋繁  
大島茂  
中嶋シゲル  
中島しげる

**onig3.txt - メモ帳**

ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)

/^\$/d  
s/自己照会/自己紹介/g  
s/事故紹介/自己紹介/g  
s/[小児大][島嶋][茂繁]/小島茂/g  
s/[中][島嶋][シシ][ゲゲ][ルル]/小島茂/g

**コマンドプロンプト**

```
J:\>onigsed -f onig3.txt test2.txt
```

自己紹介  
自己紹介  
小島茂  
小島茂  
小島茂  
小島茂  
小島茂  
小島茂

メタ文字の【 \* 】は直前文字に  
0回以上繰り返された文字列にマッチ  
/Go\*gle/

マッチする

0回以上にマッチ → Ggle  
1回以上にマッチ → Gogle

マッチしない

Gpgle → ×  
Gopgle → ×

- 出典元  
<http://www.rubylife.jp/regexp/repeat/index2.html>



# メタ文字の【 + 】は 直前文字に1回以上マッチ

/Go+gle/

マッチする

Gogle

Google

マッチしない

Ggle

Gopgle

- 出典元  
<http://www.rubylife.jp/regexp/repeat/index3.html>

メタ文字の【 . 】は  
改行を除く、任意の一文字にマッチ

/ab.cd/

マッチする

abhcd

ab4cd

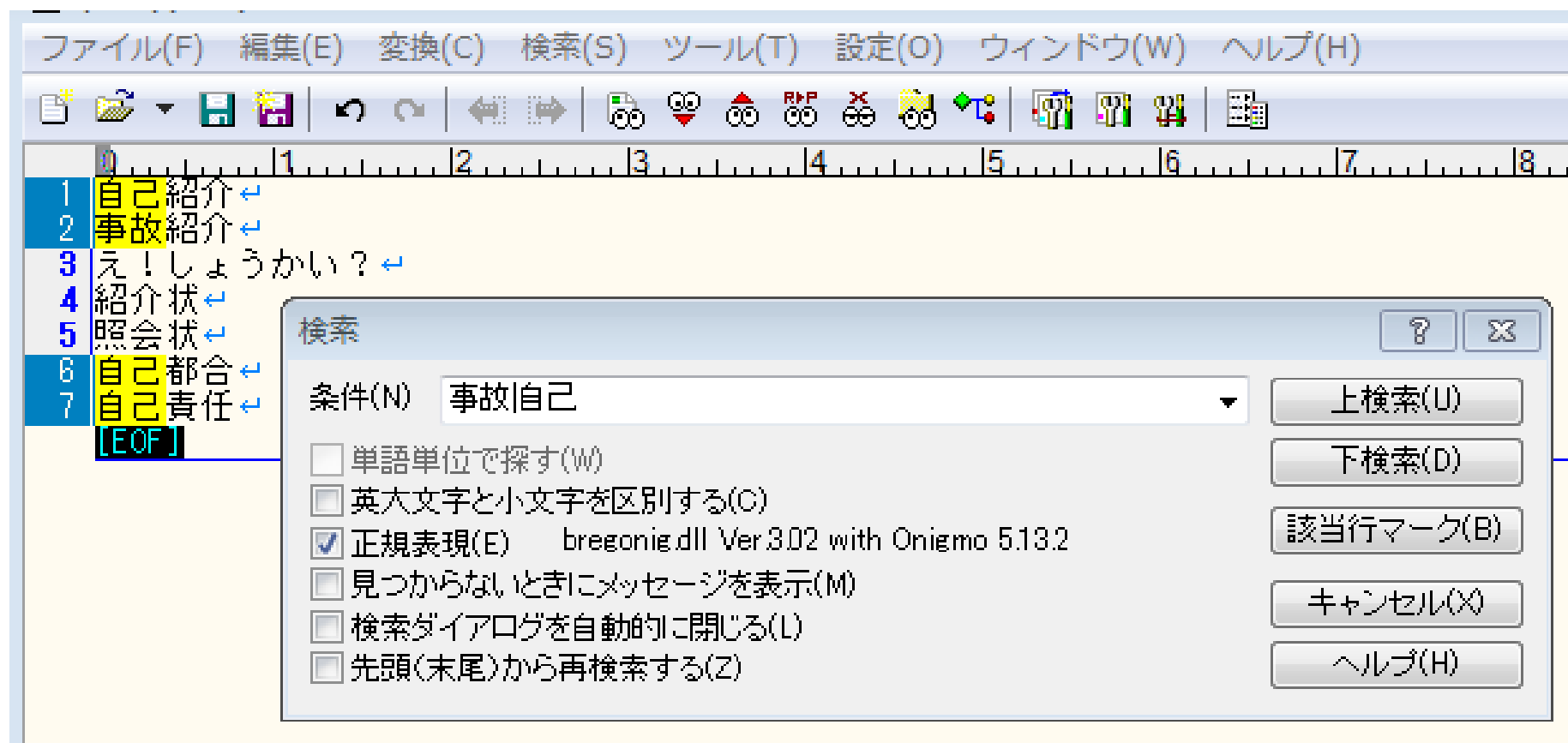
マッチしない

abcd

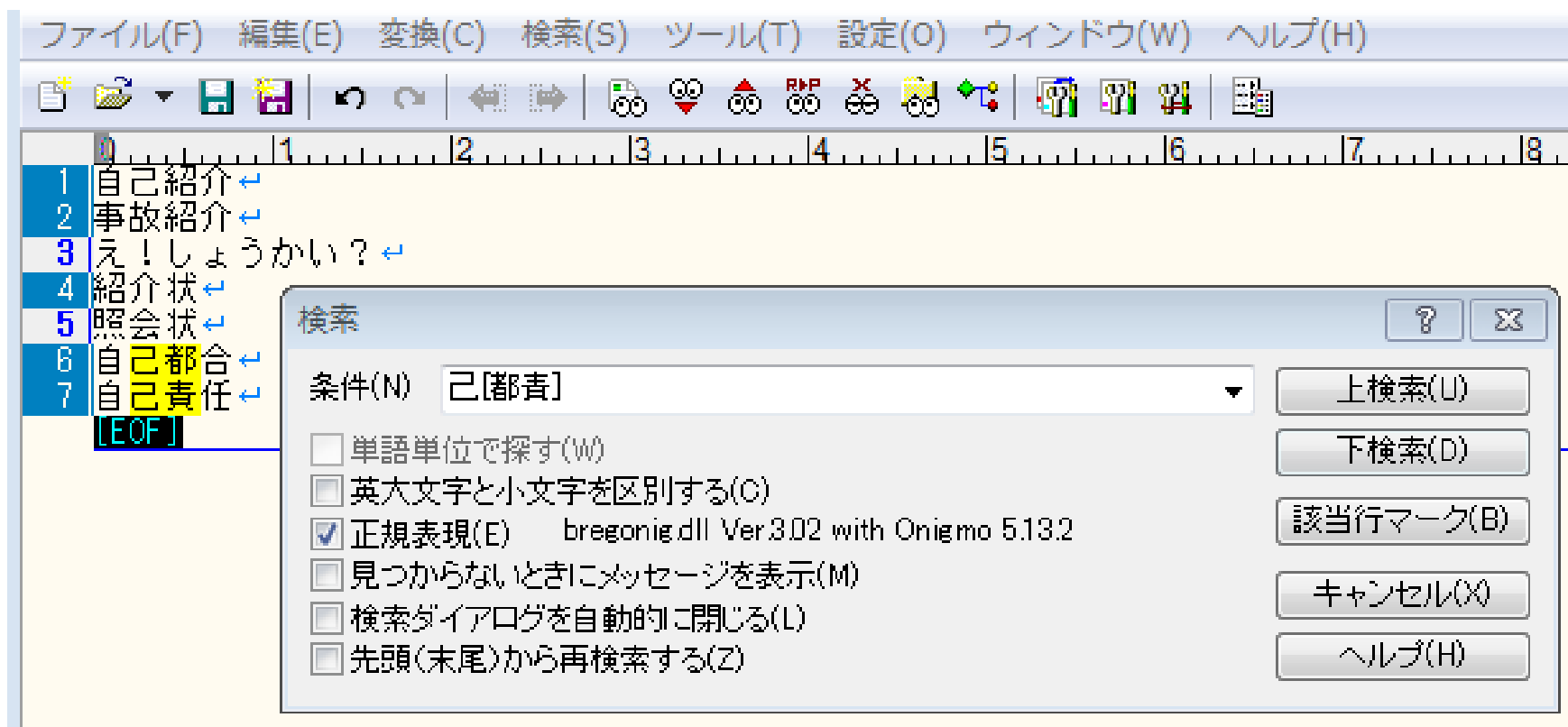
abppcd

- 出典元  
<http://www.rubylife.jp/regexp/repeat/index1.html>

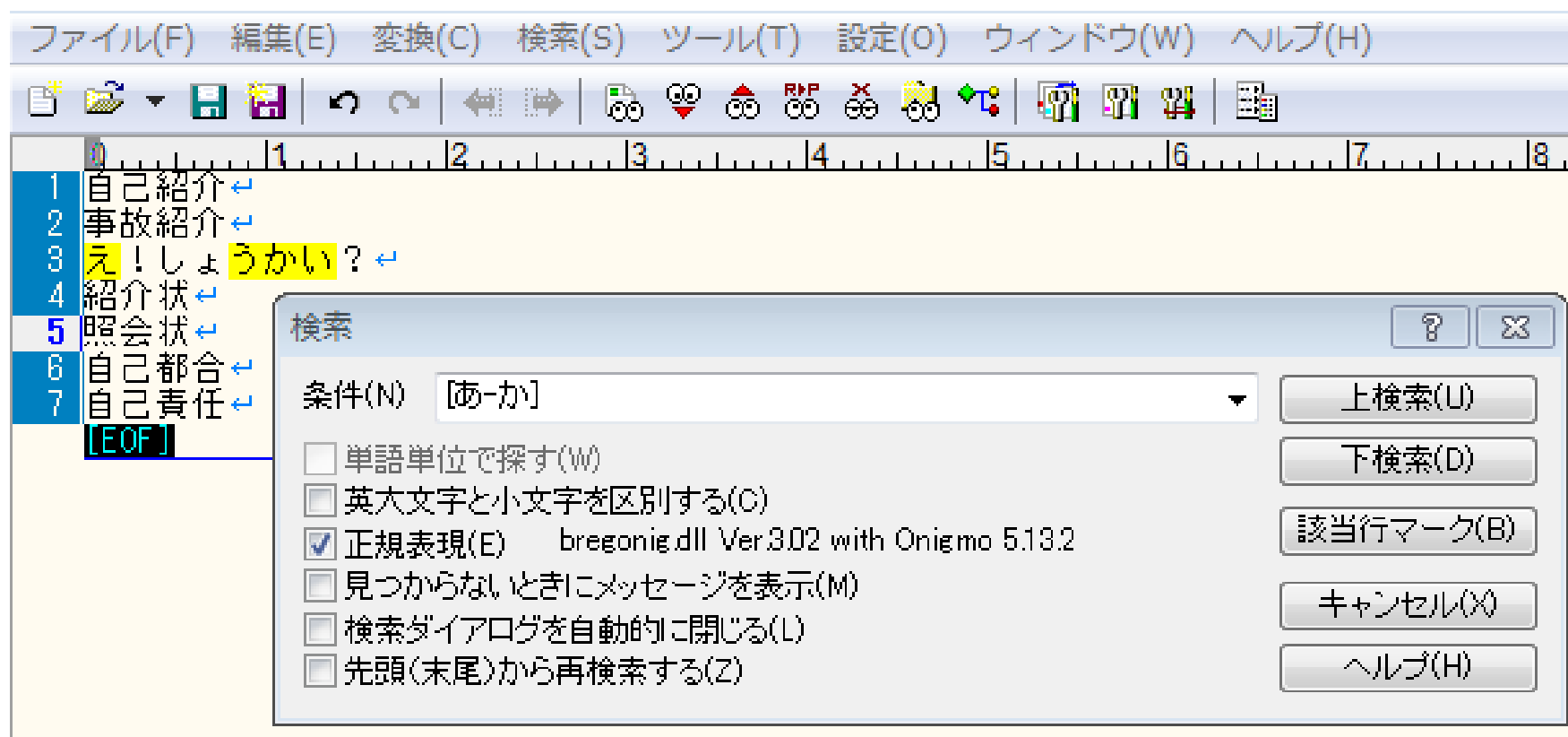
# 検索： 事故|自己（どちらか）



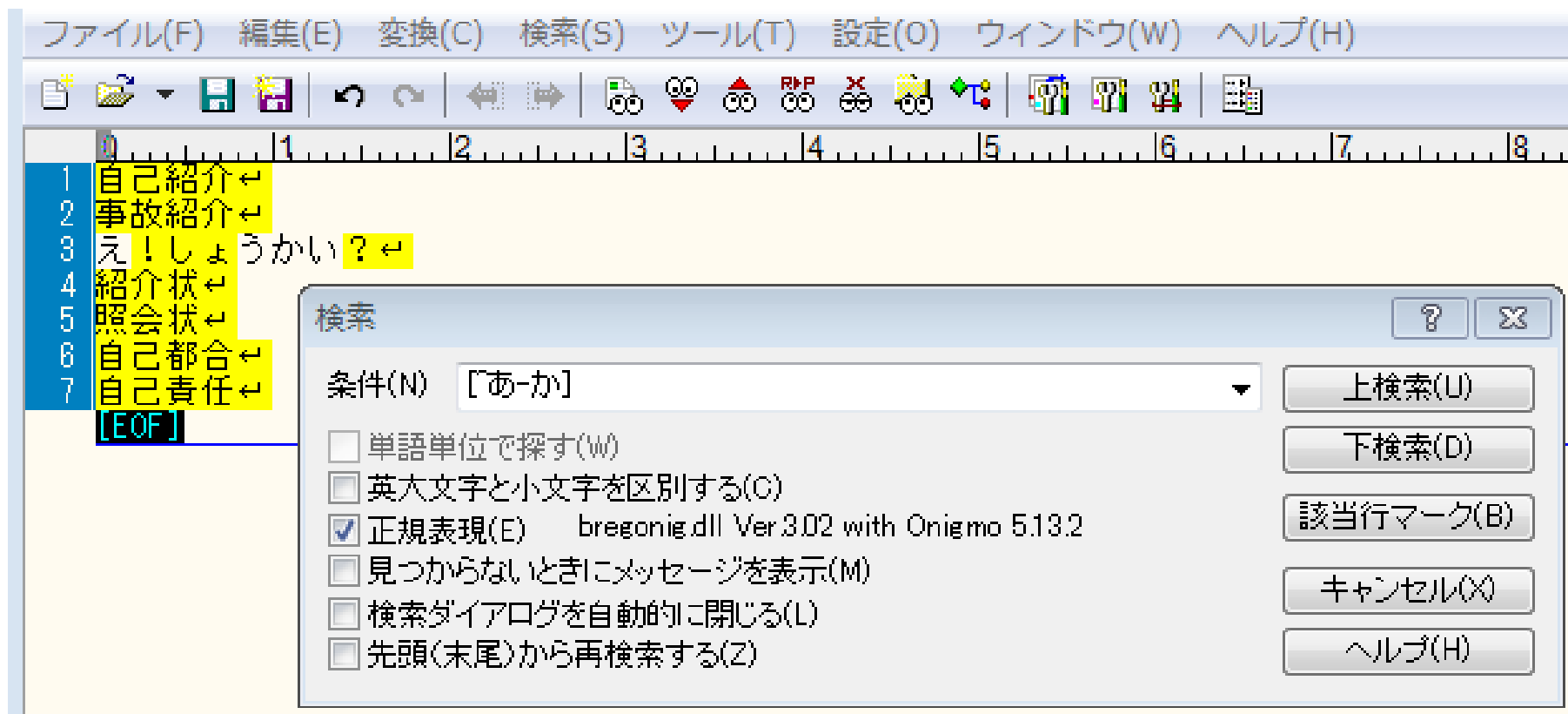
# 検索: 己[都責] 己と( 都 or 責)



# 検索： [あ-か] （あ～か迄）

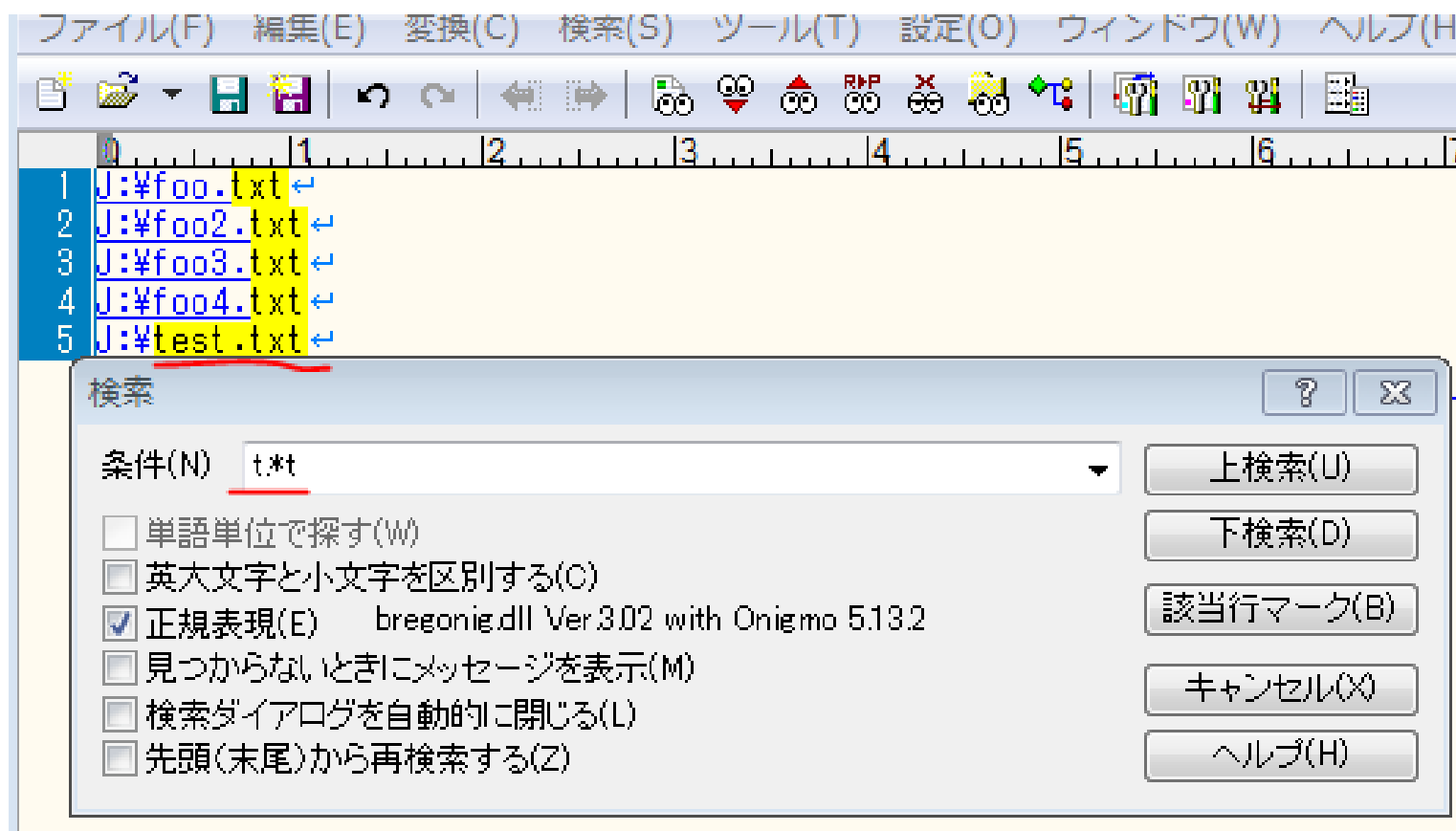


# 検索： [^あ-か]（あ～か迄 以外）



# 正規表現の基本は【最長一致】

`t.*t` → `test.txt` (最後の`t`にマッチ)



# 基本動作【最長一致】

## 最後の 8 にマッチ

1					
2	67626:2015/11/26	00:28	27,046	DropboxUpdate.log-2015-11-25-15-28-00-193-9564-finished	↵
3	67627:2015/11/26	00:28	49,678	DropboxUpdate.log-2015-11-25-15-28-17-400-9884-finished	↵
4	67628:2015/11/26	00:28	3,526	DropboxUpdate.log-2015-11-25-15-28-17-446-9564-finished	↵
5	67629:2015/11/26	01:28	22,818	DropboxUpdate.log-2015-11-25-16-28-00-197-8084-finished	↵
6	67630:2015/11/26	02:28	20,594	DropboxUpdate.log-2015-11-25-17-28-00-120-448-finished	↵
7	67631:2015/11/26	03:28	20,762	DropboxUpdate.log-2015-11-25-18-28-00-330-1612-finished	↵
8	67632:2015/11/26	04:28	20,914	DropboxUpdate.log-2015-11-25-19-28-00-223-10188-finished	↵
9	67633:2015/11/26	05:28	27,046	DropboxUpdate.log-2015-11-25-20-28-00-268-1888-finished	↵
10	67634:2015/11/26	05:28	49,678	DropboxUpdate.log-2015-11-25-20-28-46-974-7152-finished	↵
11	67635:2015/11/26	05:28	3,526	DropboxUpdate.log-2015-11-25-20-28-47-021-1888-finished	↵
12	67636:2015/11/26	06:28	22,818	DropboxUpdate.log-2015-11-25-21-28-00-123-5448-finished	↵
13	67637:2015/11/26	07:28	20,762	DropboxUpdate.log-2015-11-25-22-28-00-276-8600-finished	↵
14	67638:2015/11/26	08:28	20,762	DropboxUpdate.log-2015-11-25-23-28-00-186-5572-finished	↵
15	67639:2015/11/26	09:28	20,762	DropboxUpdate.log-2015-11-26-00-28-00-068-6444-finished	↵
16	67640:2015/11/26	10:28	27,278	DropboxUpdate.log-2015-11-26-01-28-00-300-5332-finished	↵
17	67641:2015/11/26	10:28	50,006	DropboxUpdate.log-2015-11-26-01-28-57-084-4076-finished	↵
18	67642:2015/11/26	10:28	3,566	DropboxUpdate.log-2015-11-26-01-28-57-115-5332-finished	↵
19	67643:2015/11/26	11:28	22,818	DropboxUpdate.log-2015-11-26-02-28-00-270-9632-finished	↵
20	67644:2015/11/26	12:28	20,762	DropboxUpdate.log-2015-11-26-03-28-00-285-8148-finished	[EOF]

置換

置換前(N)

置換後(P)

☐ クリップボードから貼り付ける(T)  
☐ 単語単位で探す(W)  
☐ 英大文字と小文字を区別する(O)  
☒ 正規表現(E)  
breonig.dll Ver.3.02 with Onigmo 5.13.2  
☐ 「すべて置換」は置換の繰返し(I)  
☐ 見つからないときにメッセージを表示(M)  
☐ 置換ダイアログを自動的に閉じる(L)  
☐ 先頭(末尾)から再検索する(Z)

置換対象

☒ 選択文字(O)  
☐ 選択始点(1)挿入  
☐ 選択終点(2)追加

範囲

☐ 選択範囲(S)  
☒ ファイル全体(O)

上検索(U)

下検索(D)

該当行マーク(B)

置換(R)

すべて置換(A)

キャンセル(X)

ヘルプ(H)



# 【最短一致】は → ? を付加 最初の 8 にマッチ

1			
2	87626:2015/11/26 00:28	27,046	DropboxUpdate.log-2015-11-25-15-28-00-193-9564-finished
3	87627:2015/11/26 00:28	49,678	DropboxUpdate.log-2015-11-25-15-28-17-400-9884-finished
4	87628:2015/11/26 00:28	3,526	DropboxUpdate.log-2015-11-25-15-28-17-446-9564-finished
5	87629:2015/11/26 01:28	22,818	DropboxUpdate.log-2015-11-25-16-28-00-197-6084-finished
6	87630:2015/11/26 02:28	20,594	DropboxUpdate.log-2015-11-25-17-28-00-120-448-finished
7	87631:2015/11/26 03:28	20,762	DropboxUpdate.log-2015-11-25-18-28-00-330-1612-finished
8	87632:2015/11/26 04:28	20,914	DropboxUpdate.log-2015-11-25-19-28-00-223-10188-finished
9	87633:2015/11/26 05:28	27,046	DropboxUpdate.log-2015-11-25-20-28-00-268-1888-finished
10	87634:2015/11/26 05:28	49,678	DropboxUpdate.log-2015-11-25-20-28-46-974-7152-finished
11	87635:2015/11/26 05:28	3,526	DropboxUpdate.log-2015-11-25-20-28-47-021-1888-finished
12	87636:2015/11/26 06:28	22,818	DropboxUpdate.log-2015-11-25-21-28-00-123-5448-finished
13	87637:2015/11/26 07:28	20,762	DropboxUpdate.log-2015-11-25-22-28-00-276-8600-finished
14	87638:2015/11/26 08:28	20,762	DropboxUpdate.log-2015-11-25-23-28-00-186-5572-finished
15	87639:2015/11/26 09:28	20,762	DropboxUpdate.log-2015-11-26-00-28-00-068-6444-finished
16	87640:2015/11/26 10:28	27,278	DropboxUpdate.log-2015-11-26-01-28-00-300-5332-finished
17	87641:2015/11/26 10:28	50,006	DropboxUpdate.log-2015-11-26-01-28-57-084-4076-finished
18	87642:2015/11/26 10:28	3,566	DropboxUpdate.log-2015-11-26-01-28-57-115-5332-finished
19	87643:2015/11/26 11:28	22,818	DropboxUpdate.log-2015-11-26-02-28-00-270-9632-finished
20	87644:2015/11/26 12:28	20,762	DropboxUpdate.log-2015-11-26-03-28-00-285-8148-finished

置換

☐ クリップボードから貼り付ける(T)  
☐ 単語単位で探す(W)  
☒ 英大文字と小文字を区別する(C)  
☒ 正規表現(E)  
     bregonig.dll Ver 3.02 with Onigmo 5.13.2  
☐ 「すべて置換」は置換の繰返し(I)  
☐ 見つからないときにメッセージを表示(M)  
☐ 置換ダイアログを自動的に閉じる(L)  
☐ 先頭(末尾)から再検索する(Z)

置換対象

☒ 選択文字(O)  
☐ 選択始点(1)挿入  
☐ 選択終点(2)追加

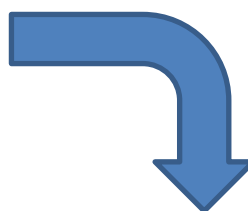
範囲

☐ 選択範囲(S)  
☒ ファイル全体(O)

# 正規表現を使いフォーマット変換(置換)

ファイル(F) 編集(E) 変換(C)

1 "1991年,A,値5" ←  
2 "1992年,B,値15" ←  
3 "1993年,B,値555" ←  
4 "1994年,C,値54321" ←  
5 "1995年,C,値5987654" ←  
[EOF]



ファイル(F) 編集(E) 変換(C) 検索(S) ツール(T)

1 {"year": 1991, "name": "A", "value": 5} ←  
2 {"year": 1992, "name": "B", "value": 15} ←  
3 {"year": 1993, "name": "B", "value": 555} ←  
4 {"year": 1994, "name": "C", "value": 54321} ←  
5 {"year": 1995, "name": "C", "value": 5987654} ←  
[EOF]

( ) → は値をグループ化  
\$1 → はグループ化1番目の値

置換

置換前(N) "(\*)年,(\*)値(\*)"

置換後(P) {"year": \$1,"name": "\$2","value": \$3}

☐ クリップボードから貼り付ける(T)  
☐ 単語単位で探す(W)  
☐ 英大文字と小文字を区別する(O)  
☒ 正規表現(E)  
bregonig.dll Ver.302 with Onigmo 5.13.2

☐ 「すべて置換」は置換の繰返し(I)  
☐ 見つからないときにメッセージを表示(M)  
☐ 置換ダイアログを自動的に開じる(L)  
☐ 先頭(末尾)から再検索する(Z)

置換対象

☒ 選択文字(O)  
☐ 選択始点(1)挿入  
☐ 選択終点(2)追加

範囲

☐ 選択範囲(S)  
☒ ファイル全体(O)

上検索(U)  
下検索(D)  
該当行マーク(B)  
置換(R)  
すべて置換(A)  
キャンセル(X)  
ヘルプ(H)

【実現したいこと】 青文字部分だけを取り出して → フォーマットを変える

"1991 年,A,値 5" というデータを → {"year": 1991, "name": "A", "value": 5} にしたい

【置換前】 取り出したい青文字部分を ( ) で囲みグループ化する (赤文字がデータ区切り記号になる)

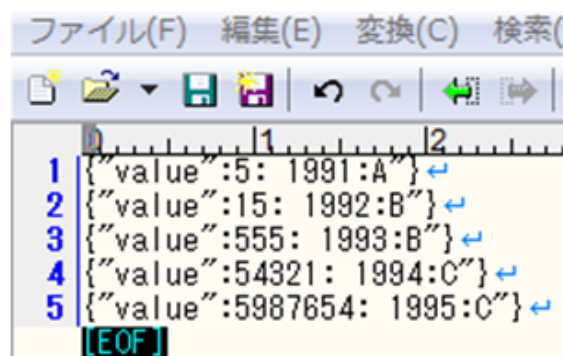
"(.\*)年,(.\*)値(.\*)"

【置換後】 グループ化した青文字データを使い新しいフォーマットに変更する

{"year": \$1,"name": "\$2","value": \$3}

【応用例】

{"value": \$3: \$1: \$2}

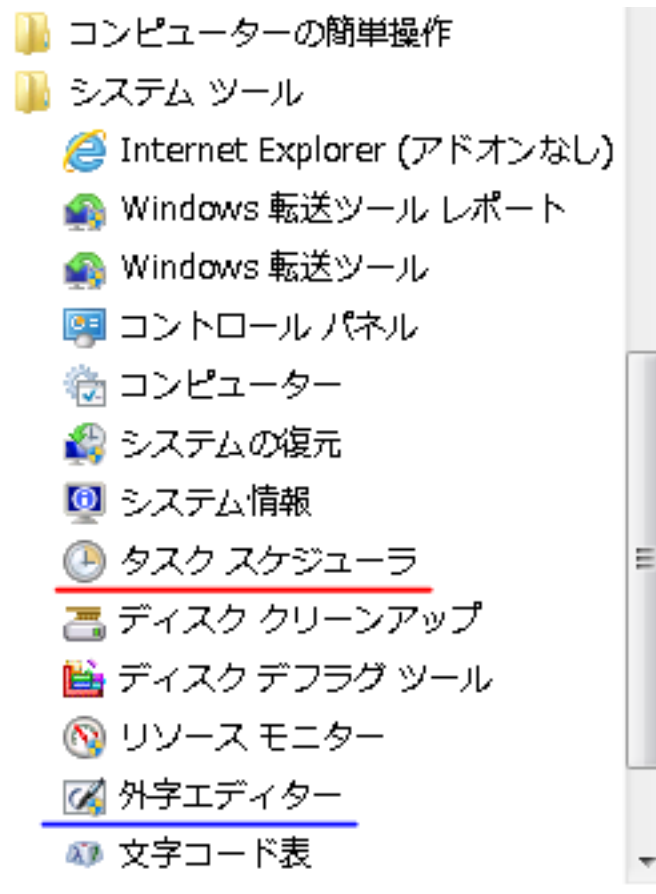
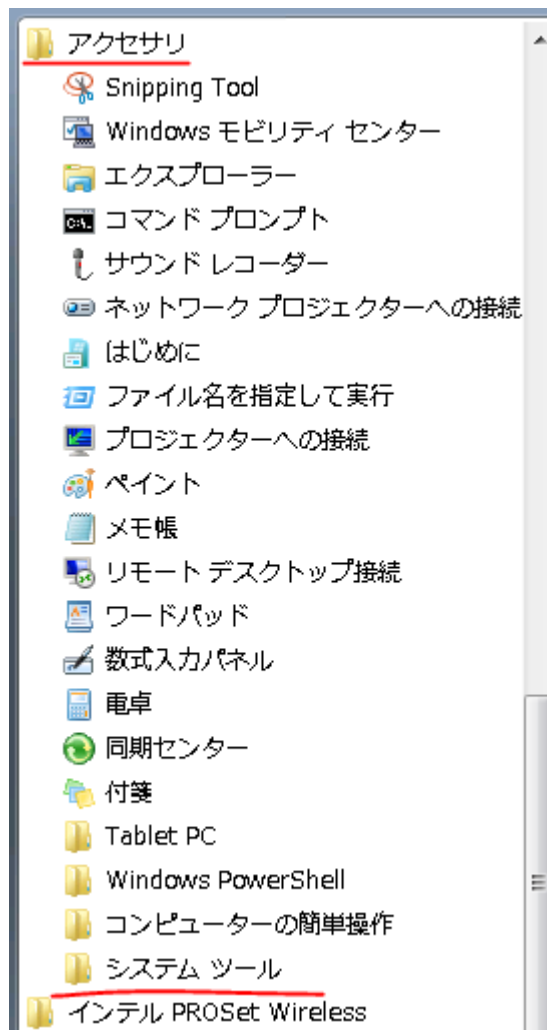


以上

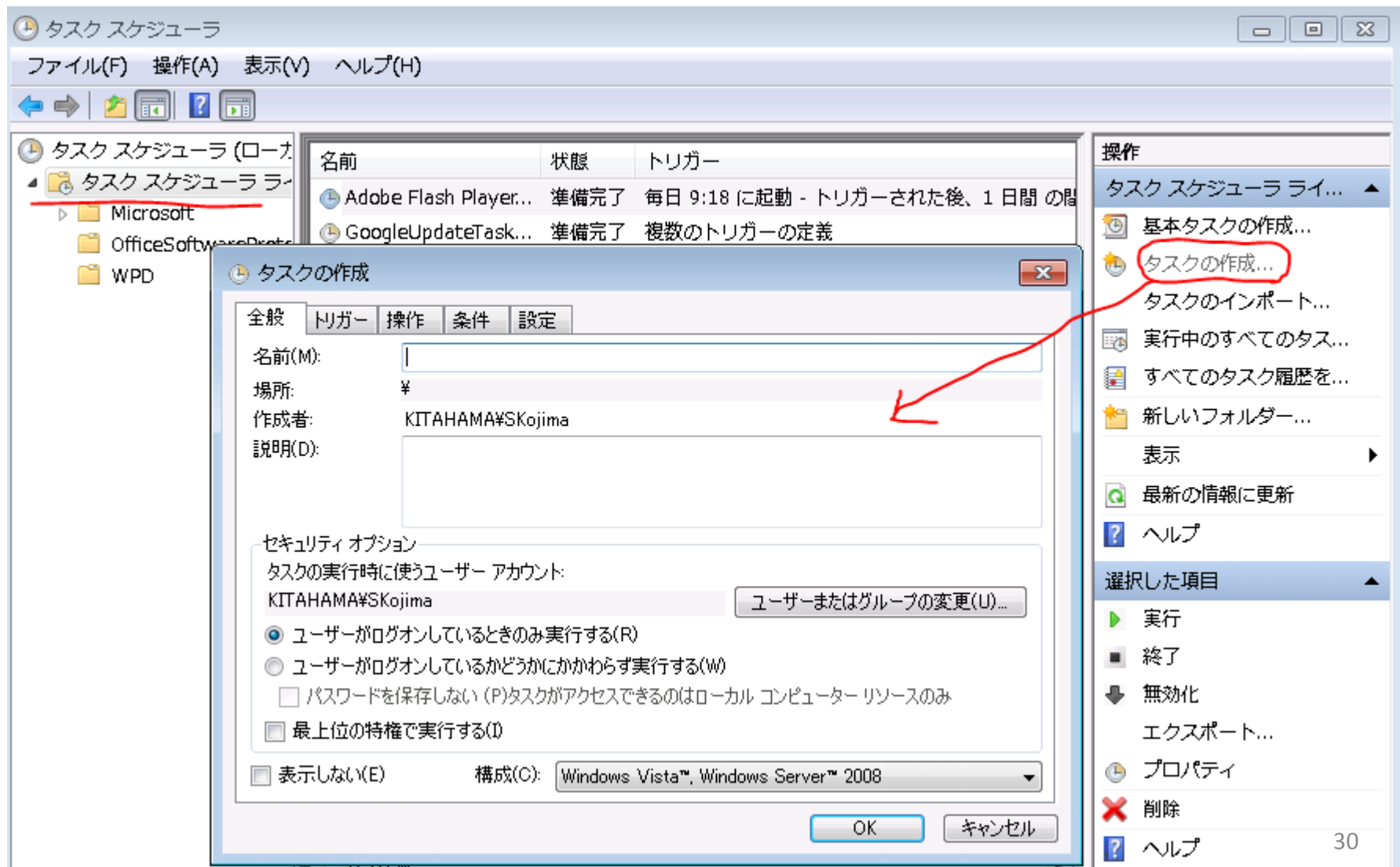
# おまけ

# 定型業務を自動化するには？

## 【タスク スケジューラ】で定期的に行



# タスクの作成



# 実行ユーザ名とパスワードを指定

fnsのプロパティ (ローカル コンピューター)

全般 トリガー 操作 条件 設定 履歴

名前(M): fns

場所: ¥

作成者: KITAHAMA¥Skojima

説明(D):

セキュリティ オプション

タスクの実行時に使うユーザー アカウント:  
kitahama¥skojima ユーザーまたはグループの変更(U)...

☐ ユーザーがログオンしているときのみ実行する(R)

☒ ユーザーがログオンしているかどうかにかかわらず実行する(W)

☐ パスワードを保存しない (P)タスクがアクセスできるのはローカル コンピューター リソースのみ

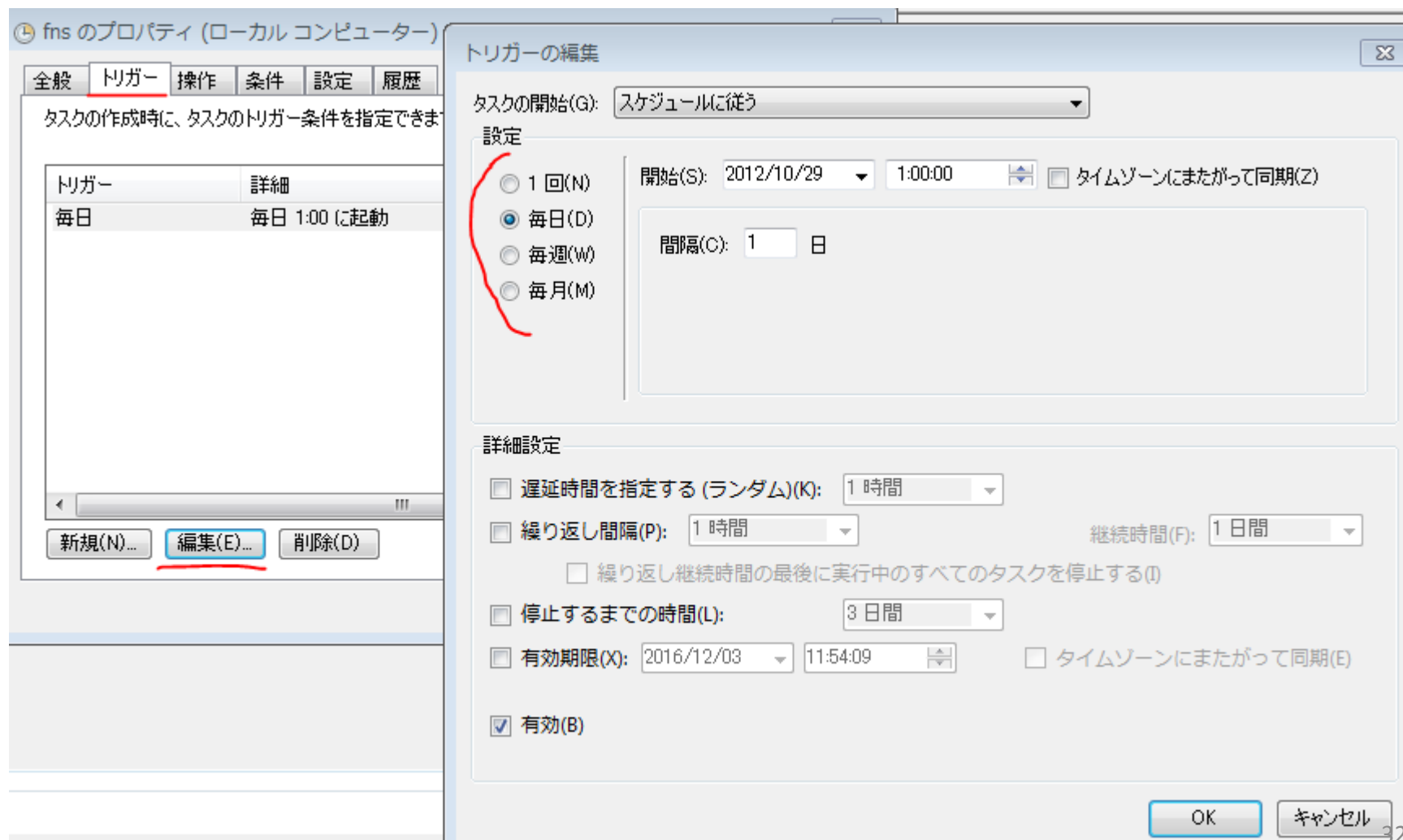
☒ 最上位の特権で実行する(I)

☐ 表示しない(E) 構成(C): Windows Vista™, Windows Server™ 2008

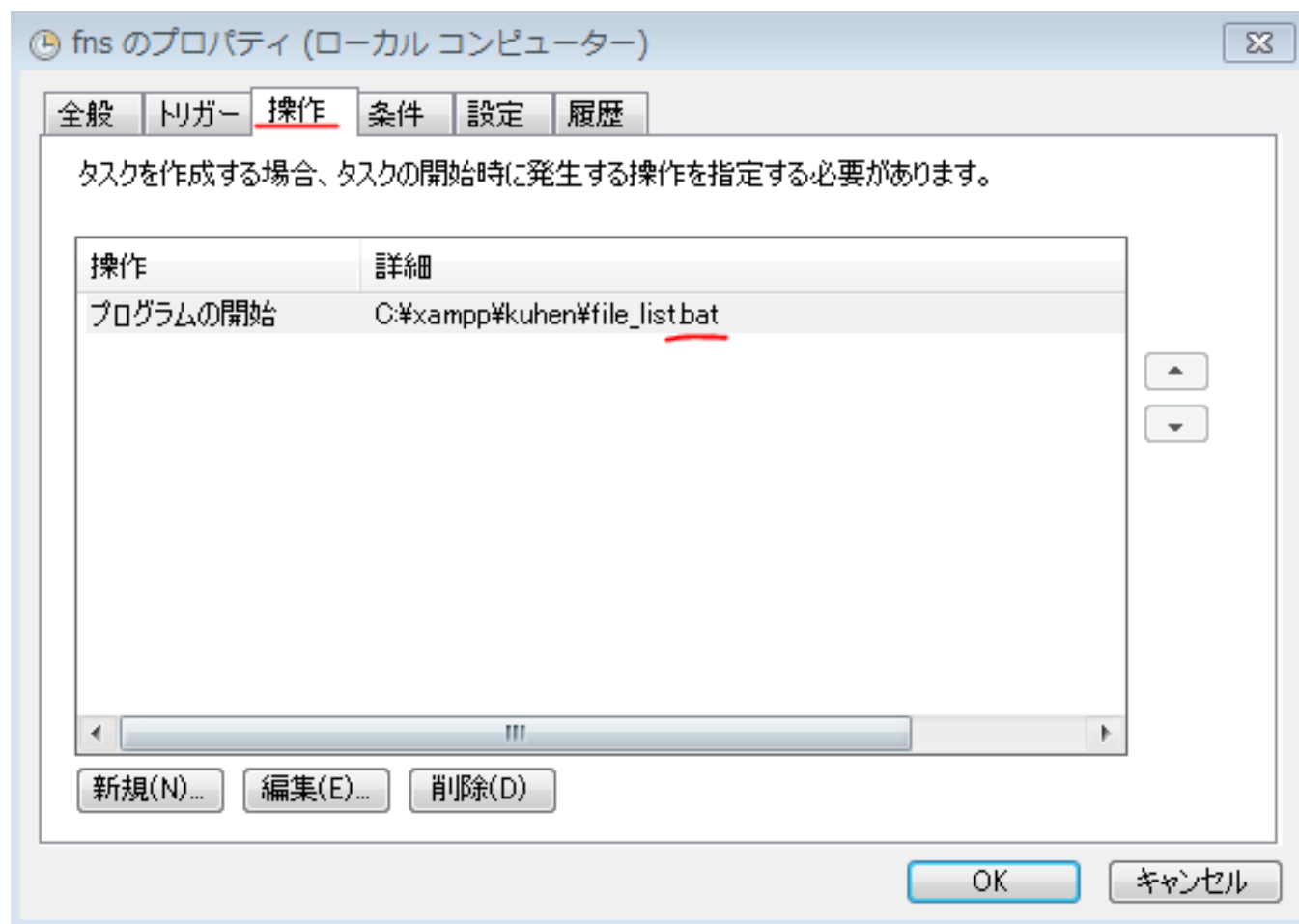
OK キャンセル



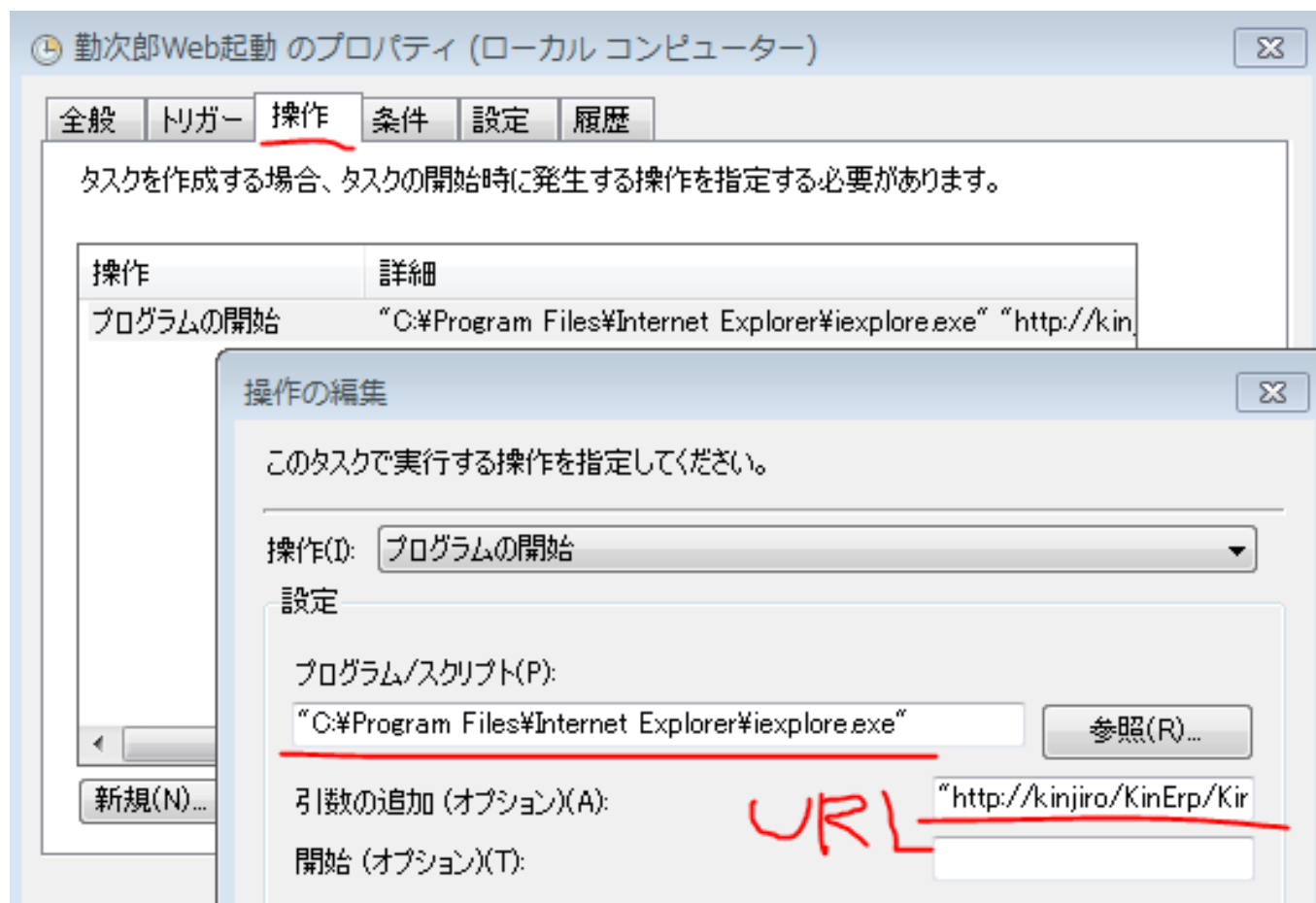
# トリガー(実行時間を指定)



# 例えば、BATファイルを実行



# 例えば、決められたURLを起動

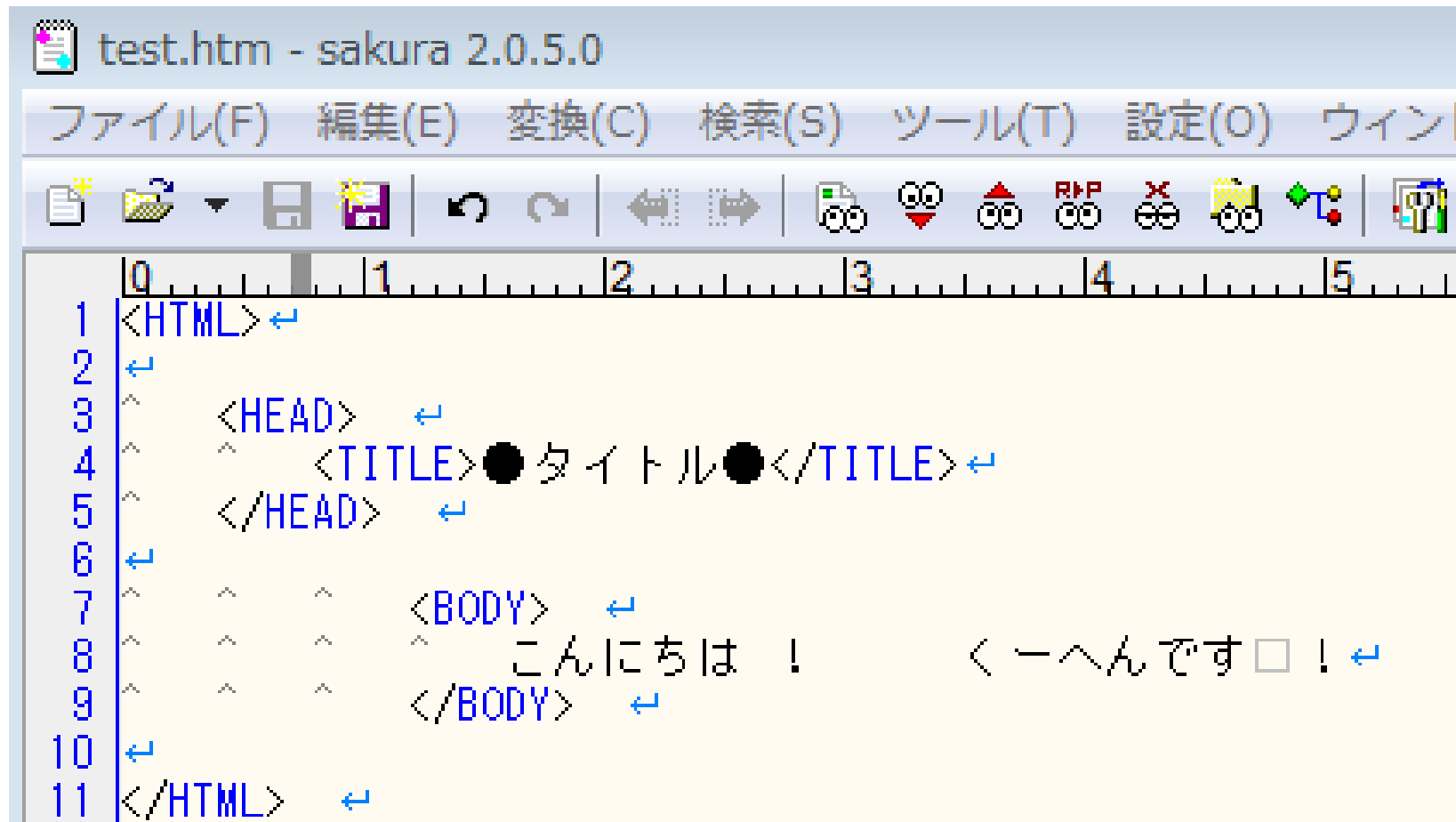


# また時間がある時に 各自でお試し下さい

- 実際に実行していくと時間が無いのでパソコン内の指定フォルダを使いその検索結果を→HTML形式で作成&表示させます
- ソースコードや、このスライドは以下のGithubに掲載中です

<https://github.com/flhtc1964/Computer/tree/master/2015-12-19>

# 最低限のHTML形式ファイルを バッチ処理で自動作成する



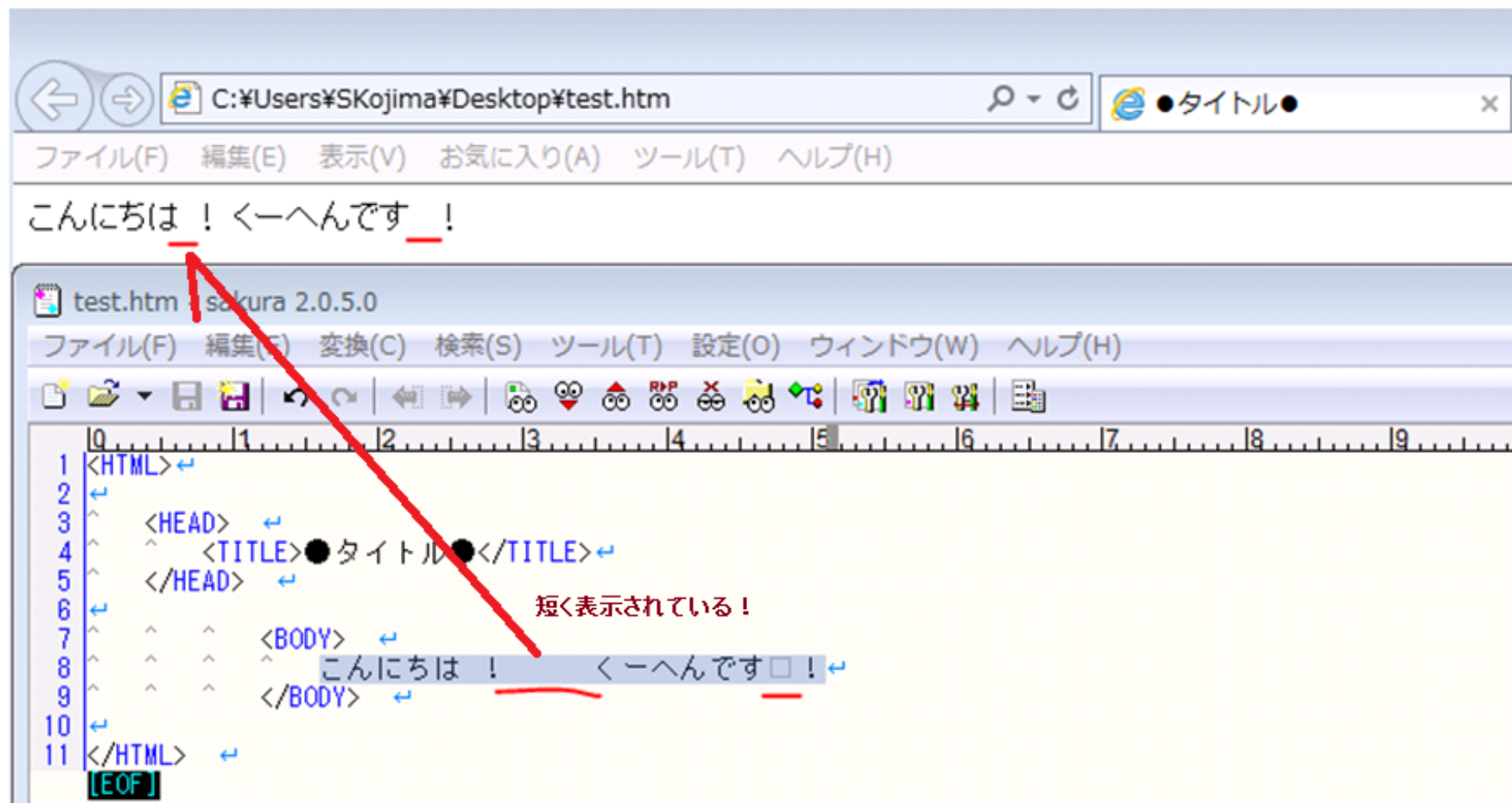
The screenshot shows the Sakura 2.0.5.0 HTML editor interface. The title bar reads "test.htm - sakura 2.0.5.0". The menu bar includes "ファイル(F)", "編集(E)", "変換(C)", "検索(S)", "ツール(T)", "設定(O)", and "ウィンドウ(W)". The toolbar contains various icons for file operations, editing, and viewing. The main text area displays a minimal HTML document structure with line numbers 1 through 11 on the left. The code is as follows:

```
1 <HTML> ←
2 ←
3 ^   <HEAD>   ←
4 ^     ^   <TITLE>●タイトル●</TITLE> ←
5 ^   </HEAD>  ←
6 ←
7 ^     ^     ^   <BODY>   ←
8 ^     ^     ^     ^   こんにちは ！       < ーへんです□！ ←
9 ^     ^     ^   </BODY> ←
10 ←
11 </HTML>  ←
```

# <HTML>で始まり </HTML>で終わる

- 取りあえずメモ帳で記述可能
- <タグ>文字列</タグ>で文字を修飾  
例: 文字の**大きさ**、文字の**色**等
- 高度になると使用する文字等指定
- <br>タグを入れないと改行せず表示される  
基本ワープロと同じく文字は連続表示
- 使用するブラウザーにより、表示や動作が微妙に違う

# 半角スペースは無視される罠



# デモを実施

- ブラウザーによってフォルダのリンクを開くことが出来ない場合がありますので、今回はIEを使ってデモをします



最後に

(自己 | 事故) 紹介

とは実は

自己紹介 or 事故紹介  
に「マッチ」を意味する！