

## K2 エディタの使い方

2005.5 小木曾智信 togiso@meikai.ac.jp

「K2 エディタ」はエディタの一種です。文章の入力などにも利用できますが、検索や文字列処理の機能が強力なのが特長です。こうした機能を使いこなすことで様々な目的に応用できます。パソコンを道具として使いこなすためにぜひエディタの使い方に慣れて欲しいと思います。

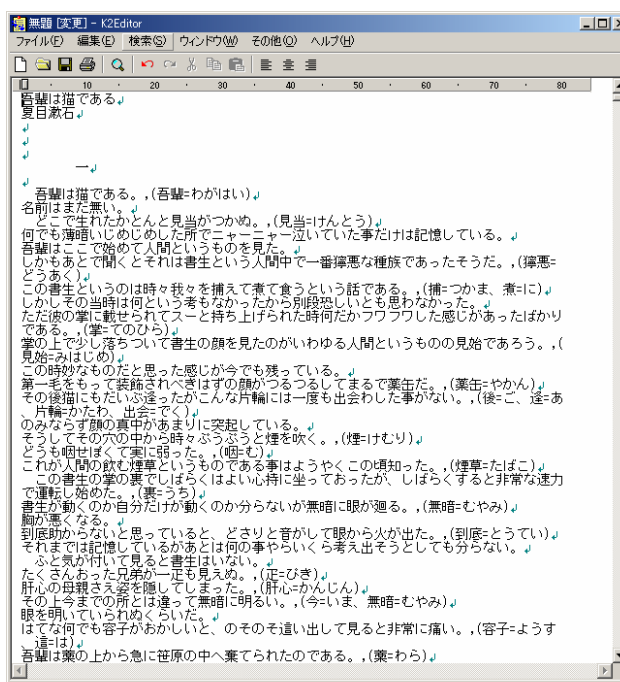
### K2 エディタの起動

K2 エディタは、「スタート」メニューの「プログラム」にあるアイコンから起動できます。

右の画面は、K2 エディタで夏目漱石の小説「吾輩は猫である」のテキストファイルを開いたところです。

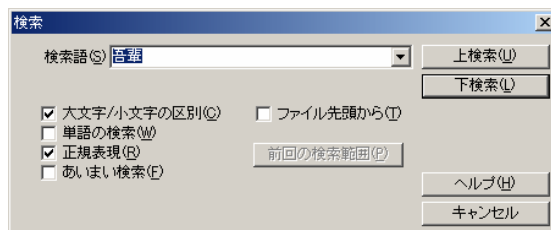
今回は検索に関連した機能に絞って説明しますが、これらの機能は、全てメニューの「検索」から呼び出します。

なお、以下の例では日本語のテキストを例にして説明していますが、英文や数値データなどでも同じような処理が可能です。自分の目的にあわせて適宜読み替えてください。



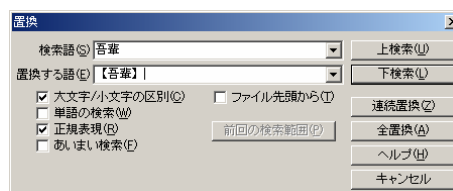
### 検索と置換

「検索」は開いているファイルの中から特定の言葉（文字列）を見つけ出してそこにジャンプする機能です。ワープロやブラウザなどにも搭載されているので馴染み深いものだと思います。K2 エディタではメニューの「検索」から「検索(F)」を選ぶと、右のようなダイアログボックスが現れ、ファイル内を検索できます（ここに出ている「正規表現」については後で詳しく説明します）。



[ F3 ] キーでカーソルがあるところより後の候補を検索、[ Shift ] キーを押しながら [ F3 ] キーで前の候補を検索できます。

「置換」は「検索」と同じように、開いているファイルの中から特定の文字列を見つけ出して、それを他の文字列に置き換える機能です。K2 エディタではメニューの「検索」から「置換(R)」を選ぶと、右のようなダイアログボックスが現れます。



ここでは「吾輩」を「【吾輩】」に置き換えています。ここで「下検索」を押すと、カーソルがある位置から見て後の方に最初に出てくる「吾輩」を一つだけ置き換えます。「ファイル先頭から」をチェックして「全置換」ボタンを押すと、ファイルの中にある全ての「吾輩」を一度に置き換えます。

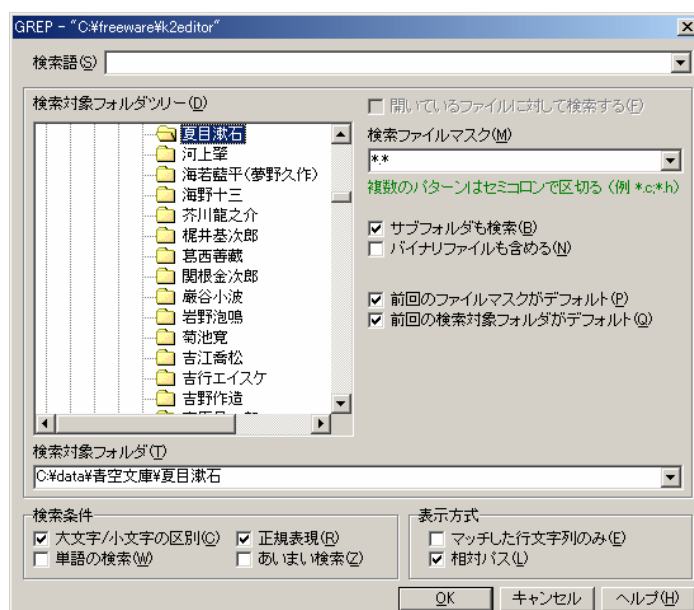
置換をうまく使うと、様々なテキスト処理に応用することができます。

## grep

「grep」とは、複数のファイルの中の特定の言葉（文字列）を含む行を検索して一覧にして取り出す機能のことです。grep では、たくさんのテキストファイルを一度に調べて一覧を作成することができます。調べたいファイルを一一つ開く必要もありません。

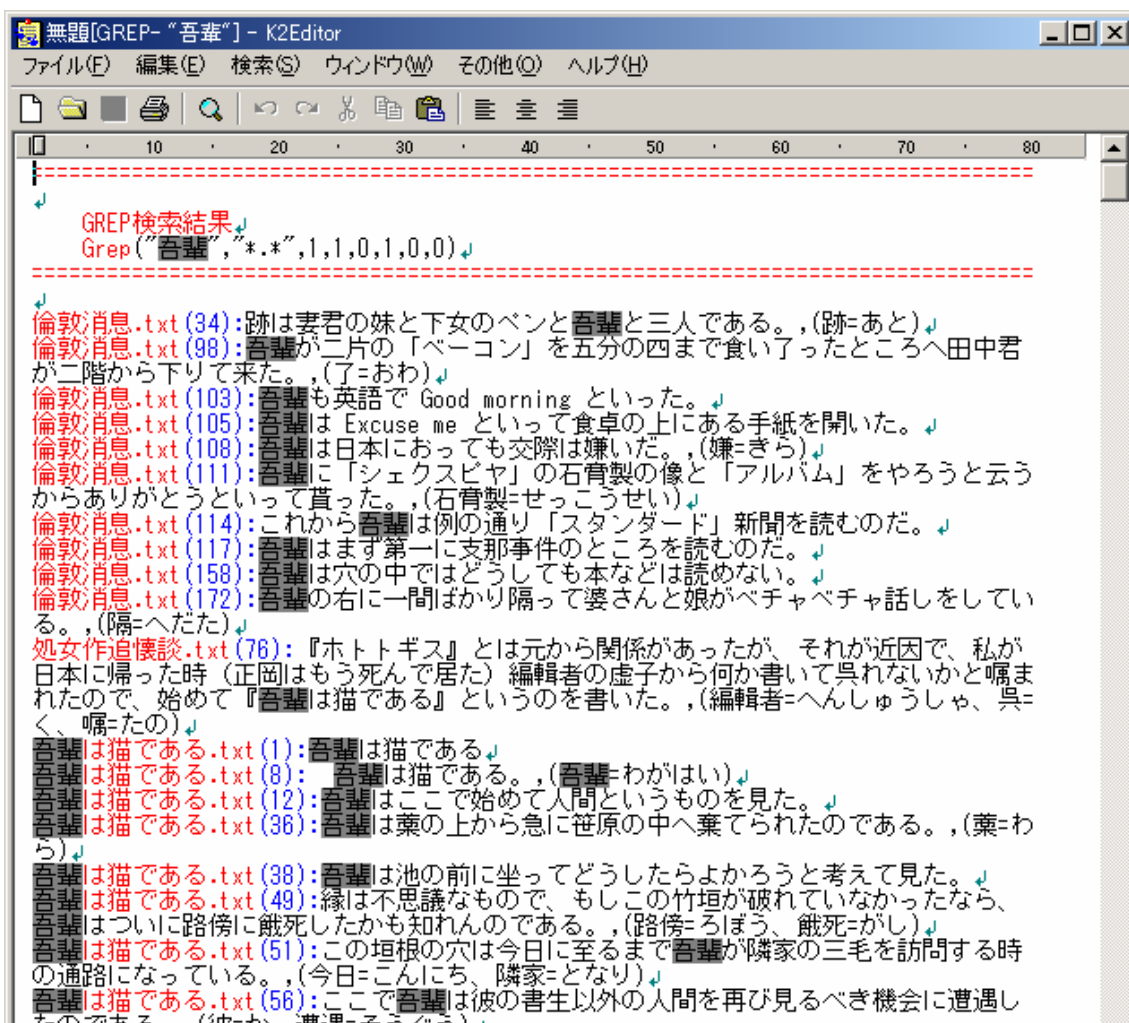
たとえば、夏目漱石の作品のなかで、「吾輩」という言葉がでてくる行を一覧にしたいときには次のようにします。（「夏目漱石」というフォルダに夏目漱石の作品のテキストファイルがたくさん入っているものとします）

- 1.メニューの「検索」から「grep」を選ぶと、次のようなダイアログボックスが現れます。
- 2.左側の「検索対象フォルダツリー」で調べたいファイルが入っているフォルダを指定します。（このとき、「サブフォルダも検索」というチェックボックスをオンにすると、「夏目漱石」フォルダの下にあるフォルダの中まで検索します。）



2. 「検索語」に調べたい言葉、ここでは「吾輩」を入力します。

3. 「OK」を押すと、新しいウィンドウがひらいて、次のような一覧が出てきます。



これで夏目漱石の作品の中で「吾輩」という言葉が使われている行の一覧ができました。

### タグジャンプ

grep の結果作られた一覧表の行頭には、次の下線部のような文字が加わっています。

吾輩は猫である.txt(12): 吾輩はここで始めて人間というものを見た。

これは、「吾輩は猫である.txt」というファイルの 12 行目に grep 検索した文字列が見つかったことを示しています。

一覧のこの行にカーソルを置いた状態で [ F10 ] キーを押すと、元のファイルが開いて当該行までジャンプしてくれます。この場合は「吾輩は猫である.txt」を自動でひらいて、12 行目にジャンプします。この機能のことを「タグジャンプ」と呼びます。元のファイルを修正したり、前後の行を見渡したりするのに便利です。

(日本語入力 (IME) がオンになっていると [ F10 ] キーを押してもタグジャンプしない場合があります。このときはいったん IME をオフにして [ F10 ] キーを押してください)

## 正規表現

今までに紹介した「検索」「置換」「grep」のダイアログボックスには、「正規表現」というオプションがありました。これをオンにすると、様々な文字を表すための記号を使って高度な検索を行うことができます。プログラミング言語などでも使われることがある記法で、少しややこしいのですが、コンピュータを使いこなしていく上で覚えておくと大変便利です。K2 エディタのヘルプにも説明がありますが、ここではよく使うものを中心に簡単に説明します。「正規表現」オプションをオンにして、正規表現を使った検索を試してみてください。なお、以下の説明で出てくる正規表現に用いられている記号は全て半角です。

### キャラクタクラス

「走る」という動詞を検索したい場合には、「走る」で検索しても、「走らない」「走った」「走ろう」などを見つけることができません。かといって「走」だけで検索したのでは「走行」「競走」「疾走」など、「走る」ではないものがたくさん引っかかってしまいます。こんな場合には正規表現の「キャラクタクラス」を利用して「走[らりるれろっ]」という検索文字列で検索をするとうまくいきます。キャラクタクラスとは、[]でいくつかの文字を囲んで、「括弧内の文字どれか」を表す正規表現です。

例：走[らりるれろっ] 「走ら」「走り」「走る」「走れ」「走る」「走っ」

他にも、次のような場合に使うと便利です。

例：渡[部辺邊邊] 「渡部」「渡辺」「渡邊」「渡邊」

例：[国國] 語 「国語」「國語」

例：dr[aiu]nk 「drink」「drank」「drunk」

### 範囲指定によるキャラクタクラス

キャラクタクラスは括弧 ([]) の中に一文字ずつ入れるほかに、範囲によって指定することもできます。範囲は文字コード表の順番になりますが、次のようによく使うものは覚えておくといいでしょう。

[あ-ん]	ひらがな 1 文字
[ア-ヴ]	カタカナ 1 文字
[亜-熙]	漢字 1 文字 (Shift JIS の場合)
[0-9]	数字 1 文字
[A-Za-z]	アルファベット 1 文字

これを応用して、一定範囲の数字や活用形を表すことができます。

例：198[1-5] 「1981」「1982」「1983」「1984」「1985」

例：有[ら-れ] 「有ら」「有り」「有る」「有れ」

例：明[亜-熙] 「明海」「明記」「明言」「明細」など「明」+漢字

### キャラクタクラスの補集合（否定）

括弧内の最初に「^」をいれることで、括弧内の文字以外の文字を表すことができます。

例：[^中小]辞典 「中辞典」「小辞典」以外の「～辞典」（「大辞典」など）  
範囲で指定（否定）することもできます。

[^ぁ-ん] ひらがな以外の文字

### 任意の1文字「.」

「.」で、どんな文字でも良いので何か1文字を表します。他の文字や記号と組み合わせて使います。

例：明.大学 「明海大学」「明治大学」「明星大学」「明 大学」など

### 直前の繰り返し

直前の文字や記号の繰り返しを表す正規表現として「?」「+」「\*」があります。

「?」 0回または1回の繰り返し（前の文字はあってもなくても良い）

例：明海大?生 「明海大生」または「明海生」

例：英米?語学科 「英米語学科」または「英語学科」

例：行な?う 「行なう」または「行う」

例：choo?sen? 「choose」「chose」「chosen」（「chosen」）

「+」 1回以上の繰り返し（直前の文字や記号が何個か連続する）

例：あ+っ 「あっ」「ああっ」「あああっ」「ああああっ」など

例：---+ 「---」「----」「-----」など

「\*」 0回以上の繰り返し（+と違って直前の文字はなくても良い）

例：うわあ\*っ 「うわっ」「うわあっ」「うわああっ」「うわああっ」など

### 文字列のグループ化

丸括弧と|を使って、括弧内の|で区切った文字列どれかを表すことができます。

例：(飛鳥|明日香) 「飛鳥」または「明日香」

例：(不動産|経済|外国語)学部 「不動産学部」「経済学部」「外国語学部」

### 正規表現の組み合わせ

今までに出てきたものをいくつか組み合わせた例です。

例：cho(ose|sen?) 「choose」「chose」「chosen」

例：(麻婆|マーボー?)豆腐 「麻婆豆腐」「マーボ豆腐」「マーボー豆腐」

例：暮らし?しの手[帳帖] 「暮らしの手帳」「暮しの手帳」「暮らしの手帖」「暮しの手帖」  
例：[ア-ヴ][ア-ヴ]+ カタカナが2文字以上連続する = カタカナで書かれた言葉

### エスケープシーケンス

正規表現では「.」「?」「+」「\*」「[」といった文字は今まで説明してきたように特別の意味を持っています。それでは、正規表現がオンになっているときにこれらの文字そのものを探したい場合にはどのようにしたらよいのでしょうか？ こんなときには、これらの文字の前にエスケープ文字「¥」をつけます。たとえば「?」そのものを探したい場合には「¥?」で表します。また「¥」そのものを探したい場合には「¥¥」とします。

このほか、「¥n」で「改行」、「¥t」で「タブ」という特別な記号文字をあらわすことができます。したがって「。」の後に改行を入れたい場合には、「。」を「.¥n」に置換することで実現できます。またカンマ区切りのデータの「,」を「¥t」に置換すればタブ区切りになるので、コピーして Excel にそのまま表として貼り付けることができるようになります。

### 正規表現を使った置換

今までに見てきた正規表現は、検索だけでなく置換にも使うことができます。たとえば、「行な?う」を「行う」に、「(麻婆|マーボー?)豆腐」を「麻婆豆腐」に置換して表記の統一をしたりといった具合です。

それでは、動詞「走る」の前に 印を入れるという置換、

例：「メロスが走った」 「メロスが 走った」

はどのようにすればよいでしょうか？「走[らりるれろっ]」を「 走[らりるれろっ]」に置換すると、次のような困ったことになってしまいます。

例：「メロスが走った」 「メロスが 走[らりるれろっ]た」

こういう場合にはタグ付き正規表現を使って「走([らりるれろっ])」を「 \$1」に置換することで実現できます。

### タグ付き正規表現

タグ付き正規表現を使うと、検索して一致した（置換前の）文字列をいったんっておいて、それを置換先で使うことができます。先ほどの例でいえば、「走[らりるれろっ]」に一致した「走っ」の部分をいったん保存しておいて、置換文字列「 \$1」の「\$1」の部分として後で使っています。

K2 エディタの場合には検索文字列を丸括弧「( )」でくくることで、その前後の一致した文字列を置換する文字列の中で「\$数字」という名前で利用できるようになります。括弧でくくったものが複数ある場合は前から順に「\$1」「\$2」「\$3」「\$4」...に保存されます。

タグ付き正規表現を使った置換はいろんなことに応用できます。少し例を挙げておきます。

例：電話番号の書式を書き換える（「00-0000-0000」 「00(0000)0000」）

検索：(0 [0-9] [0-9]?)-([0-9] [0-9] [0-9] ?[0-9]?)-([0-9] [0-9] [0-9] [0-9])

置換：\$1(\$2)\$3

例：ふりがなを開く（「漢字《かんじ》」 「かんじ」）

検索：[々亜-熙]+《([^ ]+ )》

置換：\$1

### 最長一致の原則

正規表現を使った置換を行う際、気をつけておかなければならない点があります。たとえばテキストファイル中に「ああああ」という文字がある場合に、「あ+」を「うう」に置換したらどうなるのかという点です。「あ+」は「あ」の連続を表しますから、「ああああ」の最初の「ああ」だけにも一致するし、「あああ」にも、「あああああ」全体にも一致することになりそうです。この置換を実際にやってみると「あああああ」全体が「うう」に変換されます（「ううあああ」などにはなりません）。正規表現は、可能な限り長い範囲に一致することになっているからです。これを最長一致の原則といいます。

この原則のために、置換が少しやっかいなことになる場合があります。次の例のようなテキストがあった場合、括弧に囲まれた任意の文字の連続を表す「『.+』」はそれぞれの『』ではなく、下線部全体に一度だけ一致してしまいます。

例：『ありがとうございます』『どういたしまして』『それではまた』

したがってタグ付き正規表現を使って『』を【】に置き換ようとして「『(.+)』」を「【\$1】」に置き換えると、

例：【ありがとうございます』『どういたしまして』『それではまた』

になってしまいます。

『』を【】に置き換えるには「『([^\s]+)』」を「【\$1】」に置換する必要があります。

（または「『(.+?)』」を「【\$1】」に置換しても同じ事になります。K2Editorでは「.+?」を使うことで最短一致になります）

例：【ありがとうございます】【どういたしまして】【それではまた】

同じようにHTMLのタグなども「<[^>]+>」または「<[^>]+?>」と表すことになります。

最後は少し難しい話になってしまいましたが、テキストエディタはパソコンを使いこなしていく上でもっとも応用範囲が広く、便利な道具の一つです。自分の目的にあわせてさまざまに工夫して使ってみてください。