

第 4 回：検索と置換

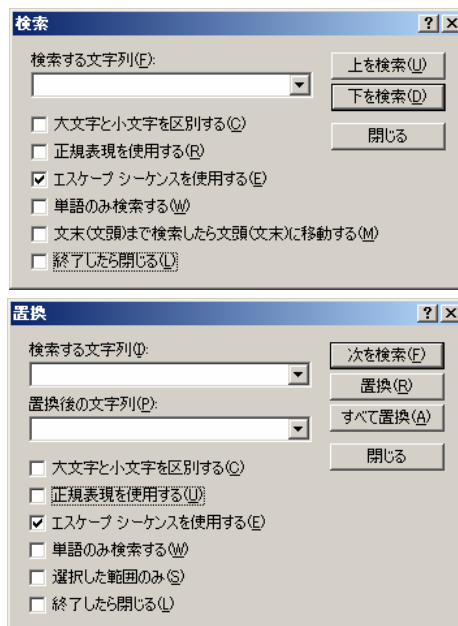
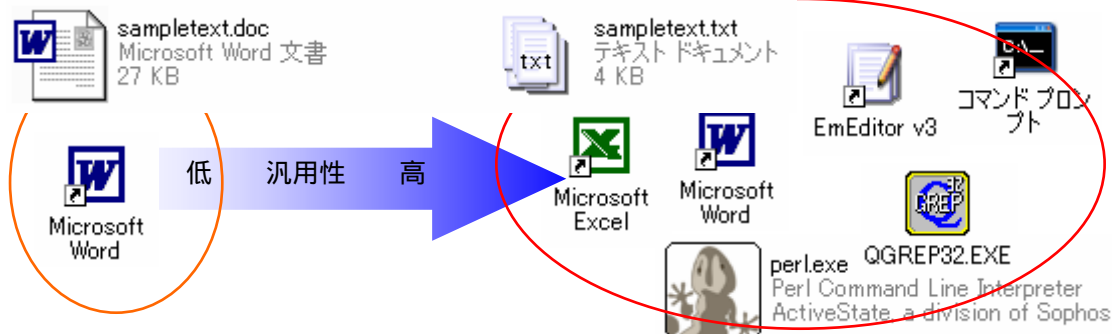
1. 検索 search と置換 replace

- デジタルデータとしてのテキストは、文字コードを頼りにテキストを網羅的に検索したり、置換したりすることができる (右図はテキストエディタ EmEditor の検索・置換ダイアログ)。
- テキストエディタの検索・置換のオプションを活用することで、複雑な検索・置換を一度におこなうことができる。

キーワード： 文字列 string, マッチ match

- さらに複雑な条件でテキストを自動整形する場合には、Perl などのスクリプト言語を利用する方法もある。この場合にも、データはテキスト形式のほうが扱いやすい。

➤ **ツールとデータの分離**: さまざまなツールで処理するテキストは汎用性の高い形式で保存しておく。



2. ソフトウェアと検索

キーワード： 検索オプション search options, エスケープシーケンス escape sequence, 正規表現 (次回取り上げる) regular expressions, あいまい検索

- Word：通常の検索と「あいまい検索」
- テキストエディタ (EmEditor Professional)：マッチした文字列のハイライト, エスケープシーケンス (表 1 参照) と正規表現 (次回取り扱う) の利用

¥b	バックスペース
¥n	改行
¥t	水平タブ
¥¥	円記号 (バックスラッシュ) 自体の検索
¥xhhhh	16 進数表記による Unicode 文字

表 1：EmEditor の主なエスケープシーケンス

- Excel：通常の検索と「オートフィルタ」による絞り込み検索
- 検索オプションの数や種類はソフトウェアにより異なる：検索したい対象がどのようなものかを考え、適切なソフトウェアを選択する。

3. 置換

- 用途：
 - 論文など作成原稿の校正
 - 研究用データの整形（「正規表現」を利用すると高度な整形も可能 次回とりあげる）

4. ファイル内・ファイル横断検索

- grep (グレップ) : 行単位でテキストを検索し, マッチした行を表示
 - grep = globally search for the regular expression and print the lines containing matches to it
- EmEditor では, [検索] [ファイルから検索] で行単位の検索を行い, 結果を抽出することができる。
 - 「正規表現」regular expression (次回扱う) を用いて複雑な検索パターンを指定し, 検索や置換をおこなうことができる。
 - さまざまなエンコード (第 6 回で扱う) のテキストの検索に対応している。¹
 - ファイル名, 行数などの情報が行頭につき, 元ファイルのリンクとして機能するので, 該当部分を EmEditor ですぐに参照することができる (元ファイルを参照したい行にカーソルを置き, **F10** キーを押すことでも参照可。この方法は, 検索結果を保存して開きなおした時など, 参照用のリンクが消えた場合でも使える)。
- ファイル名の指定 : 「ワイルドカード」wild cards
 - * = 0 文字以上の文字
 - ? = 任意の 1 文字

5. テキストエディタ (EmEditor) の利用価値

- Word, Excel よりもきめ細かい検索ができる。
 - エスケープシーケンス, 「正規表現」の利用
- EmEditor の検索でやりにくいこともある。
 - Excel や一般の検索エンジンと比較して考えてみよう

6. ファイル検索の応用 : ファイルの比較

- Excel でテキストの比較を行う
 - 2 つの列 (A 列, B 列) にそれぞれのファイルのテキストを貼り付け, C 列に以下の関数を記述 (全て半角で入力すること) :

=IF (A1=B1, "OK", "No, not the same!")

補足 : エンド・ユーザとしてのパソコン利用のこつ

- パソコンは智力・脳力をパワーアップする道具だが, すべてを代わりにやってくれるわけではない。「何をするか」「何をを使うか」「どうやってするか」を決定するのは, 利用者自身。
 - ◇ 目的をはっきりさせ, 解決の方法をよく考える。
 - ◇ 方法はひとつとは限らない。「楽に」作業を行えるよう, 手順を工夫する (「10%の努力で 70%の仕事」)。
 - ◇ すべてを知る必要はない。賢いエンド・ユーザになろう。

¹ EmEditor Professional の場合。Standard バージョンでは, Unicode テキストをファイルから検索する場合, 予めテキストを UTF-16 little endian (Unicode), BOM つき (エンコードの詳細は第 6 回授業で扱う) で保存しておくこと。