パスワードについて考えてみよう

～記号の出現確率と記号列の処理について～

＜概　要＞

アルファベットの文字から，無作為に文字を抽出してパスワードを作ることを通して，情報の形態のひとつである記号列の科学的な理解を深める。特に，記号の出現確率の意味，乱数を利用した無作為抽出の方法，記号列の処理の方法等について学ぶとともに，表計算ソフトウェアの使い方の技術も習得する。さらに，パスワードが個人情報のひとつであるという観点から，その発行や管理の仕方をシミュレーションすることにより，情報社会における安全性（セキュリティ）や個人情報の保護の大切さについて考える。

＜キーワード＞記号列，パスワード，乱数，出現確率，セキュリティ

１．学習活動

（１）導入［約２時間］

情報化社会では，様々な情報が電子化されて飛び交っている。そのような中で，大切な情報や個人的な情報を正しく管理することが重要になってくる。例えば，銀行のキャッシュカードには，持ち主だけが知っているパスワードが設定されていて，本人だけしか利用できないようになっている。もし誰かがキャッシュカードを拾って，パスワードをでたらめに想像して使っても，容易には利用できないようになっている。

さて，パスワードには英数字が使われることが多いが，でたらめにアルファベットや数字を並べるとき，それはどの程度でたらめなのであろうか。それを正しく考えるためには，その記号が出現する確率を考えなければならない。そうでないと記号列の出現する度合いについて正しい評価をすることができない。例えば町中で行き交っている自動車の番号を調べるとき，１１１１という番号を見ると珍しい番号であると感じるが，適当に並んでいる７３２８という番号は珍しいと感じない。この１１１１と７３２８という数字は，その出現確率は全く同じであるにも関わらず，人間の心理として，そのように思えないものである。ただし自動車の番号の場合は，（人間が感じる）特殊な並び，例えば１０００，８８８８，１２３４，・・などは好まれる番号なので，他の数字と比べて比較的多く出回っており，出現確率としては若干増えるかもしれない。いずれにしても記号列を考えるとき，同様の確からしさや確率分布に配慮して考えることが大切である。次にアルファベットの場合について考えてみよう。

〈課題１〉英語の文中には，アルファベット２６文字のうち，どの文字が一番使われているだろうか。また，それはどのように調べればよいだろうか。

『E が一番多く使われる。簡単な英文を印刷して，手作業で調べてみる。』

【実習１】「アルファベットの使用頻度」

①インターネット上の英語圏のサイトから，適当な英文を選ぶ。

②それを Word に取り込む。

③Word の検索機能を利用して，文中にあるアルファベットの個数を数える。

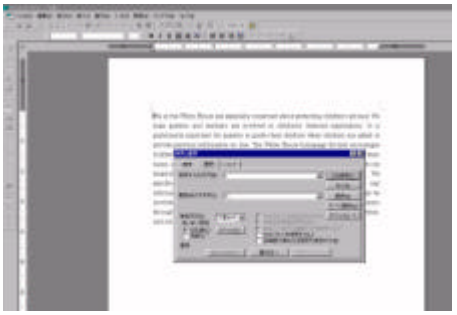

図１　Word 上での文字の検索

このように，英語という言語の中で利用されるアルファベットは，単語の成り立ちや文法の構造等により，それぞれの文字の出現確率が異なっている。アルファベット全部の文字について，その出現確率の分布状態を表したものを確率分布という。確率分布の中で起こりうるすべての事象が平等に出現すると仮定した場合，この確率分布を特に一様な確率分布という。アルファベットの文字から無作為に抽出してできた文字列を議論するときは一様分布を前提とする。

（2）乱数で遊んでみよう［約３時間］

人間の作為的な行為をできる限り排除して無作為に何かを行うために，コンピュータでは乱数というものがよく利用される。乱数とは０から９までの数字を（ある規則で）でたらめに並べたものであるが，コンピュータ上では，コンピュータ言語や各種アプリケーションソフトウェアの中で，簡単な記述や操作をすることにより自動的にそれを発生させることができる。多くの場合，それは０から１までの数字を自動発生するようになっている。

〈課題２〉インターネットで乱数についてのサイトを検索し，乱数について調べてみよう。
『例えば http://www.bunkyo.ac.jp/~nemoto/lecture/simulation/97/random/ など』

【実習２】
①Excel 上のセル C2 に，乱数を発生させる関数 RAND()を入力する。
②再計算機能（F9 キー）を利用し，何度も発生させてみる。

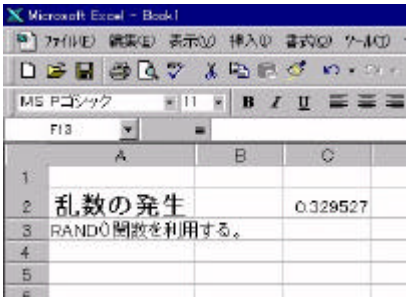

図２　Excel 上での乱数の発生

乱数は，偶然的な出来事をコンピュータ上で実現させるためによく利用させる技術であるが，ここでは Excel 上の乱数関数を使って，アルファベット A,B,C,D,E の５文字から無作為に１文字取り出すことを考えてみよう。

〈課題３〉A,B,C,D,E から無作為に１文字取り出す方法を考えてみよう。
『実習３参照』

【実習３】「無作為アルファベット発生装置」
〔方法１〕
①A,B,C,D,E を入力したセルを D5,D6,D7,D8,D9 とする。
②セル C2 に,RAND()で乱数を発生させる。
③文字を取り出すセルを C5 とし,そのセルの中に，IF 関数を使った計算式
if(+C2<0.2,+D5,if(+C2<0.4,D6,if(+c2<0.6,D7,if(+C2<0.8,D8,D9))))
を入力する。
④再計算する。
〔方法２〕
①セル C2 に,RAND()で乱数を発生させる。
②発生させた乱数を１から５までの自然数に変換するために，セル C13 に，
INT(5*RAND())+1
を入力する。
③文字を取り出すセルを F12 とし，そのセルの中に，文字列を処理する関数 MID を使った計算式
MID(“ABCDE”,+C12,1)
を入力する。
④再計算する。

（3）パスワードと出現確率［約３時間］

パスワードは，コンピュータ等に個人認証をさせるための記号列である。通常英数字を８文字程度並べたものであるが，本人にしかわからないようにしておくべきものである。従って，（人間が感じる）特殊な記号列は避けるべきであるが，本人が忘れてしまわない程度のものにすることも必要となる。

さて，アルファベット２６文字から無作為に６文字を取り出し並べてパスワードを作るとき，その安全性について考えてみよう。

〈課題４〉何通りのパスワードができるだろ

か。
『26×26×26×26×26×26 = 308,915,776 通り（Windows のアクセサリの中にある関数電卓を利用するのもよい。）』

このように，無作為に６文字を取り出し並べたパスワードは全部で約３億通りあるから，一つのパスワードの出現確率は約３億分の１ということになる。この確率から，任意の６文字パスワードを言い当てることはほとんど不可能であり，実用上は安全であるということができるかもしれない。またこの数字から，６文字パスワードで識別できる人数が約３億人であるということでもある。実社会で使われている情報システムでは，同一のパスワードが生じないように，その発行にあたっては厳格な審査の末発行される。

〈課題５〉実習２で作成した「無作為アルファベット発生装置」を利用して，A,B,C,D,E をでたらめに並べた文字列をつくるにはどうすればよいだろうか考えてみよう。

【実習４】「無作為文字列発生装置」
①セル B1 に,RAND()で乱数を発生させる。
②セル B3 に INT(5*B1)+1 を入れる。
③セル B5 に MID("ABCDE",B3,1)を入れる。
④セル B7 に B7&B5 を入れ文字列を作る。（循環参照となるため，ツールのオプションで,反復計算回数を１にしておく）
⑤セル B10 に B10+1 を入れ回数を数える。
⑥再計算機能（f9 キー）で再計算させる。

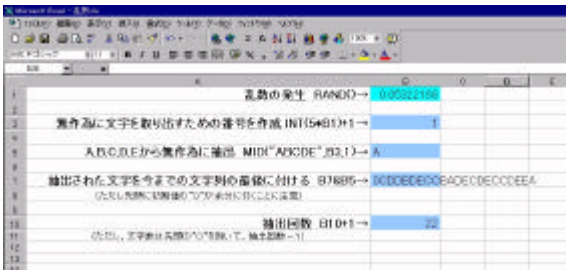

図３　Excel 上で無作為文字列を生成

（４）パスワードの発行と管理のシミュレーションをしてみよう［約２時間］

実際の情報システムにおいて，パスワードが個人識別のための重要な情報であるから，重複して発行されないように厳重に管理されている。ここで，新たに発行する必要となったパスワードが，既に使われているパスワードと一致していないかどうかを調べることを想定して，シミュレーションをしてみよう。ここでは簡単のために，アルファベットの A,B,C,D,E の５文字のみを使ってパスワードを作成し，それを管理することとする。

〈課題６〉５文字のみでは，何通りのパスワードができるだろうか。
『5×5×5×5×5 = 3,125 通り』

シミュレーションの手順は次のとおりである。
ア：数十種類のパスワードを作成し，それを発行済みパスワードとし,表を作る。
イ：この表を見やすく整理するために，アルファベット順にソートする。
ウ：任意のパスワードを一つ作る。
エ：このパスワードが今までの発行済み表にあるか否かを検索する。
オ：もしあれば発行しない。もしなければ発行済み表に加える。

この発行済みパスワードのように，ある一連の情報の集まりをデータベースと言う。コンピュータでデータベースを扱うには，専用に開発されたデータベースソフトウェアを利用することが多いが，表計算ソフトウェアにも簡単なデータベース処理機能を持っているものもある。ここでは，表計算ソフトウェアを利用して上記のシミュレーションをしてみよう。

【実習５】「Excel でのシミュレーション」
①A,B,C,D,E から無作為に文字を並べてパスワードを数十個作り，それらを発行済みパスワードとしてセルに入れる。（実習３で作成した無作為文字列発生装置を使って１００文字程度の長い文字列を作り，それを５文字ずつに切り分け,コピー・張り付け操作を繰り返し，作成してもよい。）
②この表を見やすくするために，Excel の並べ替え機能を使ってソートする。

③実習３の「無作為文字列発生装置」を使って，新しいパスワードをひとつ生成する。これを発行依頼パスワードとする。

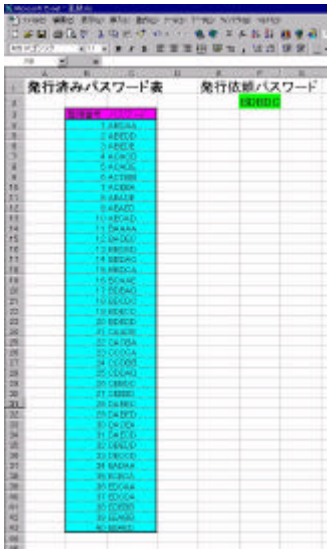

図４　Excel 上でのシミュレーション（１）

④発行済み表に対し，Excel のフィルタ機能を利用して，③のパスワードがその中にあるかどうか検索する。
⑤なかった場合は，そのパスワードを発行済み表に追加する。（発行する。）

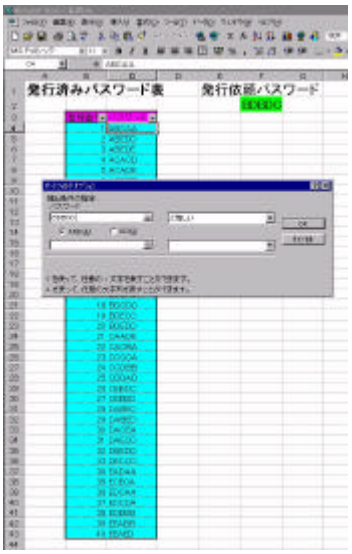

図５　Excel 上でのシミュレーション（２）

（５）情報の保護について［約２時間］

パスワードは個人情報のひとつであり，情報化社会の中で個人の権利や財産を守る大切な情報であり，その管理には細心の注意が必要である。今後の情報化社会においては，様々な情報が電子化され，情報通信ネットワークを通して流通し合う。それらの情報の中には個人情報も含まれる可能性がある。このような個人情報は不当に使われないように，法律や条令によって保護されなければならない。

〈課題７〉一般に個人情報といわれているものには，どのようなものがあるだろうか。またそれらを保護するために，どのような配慮がなされているだろうか。

情報化社会の中で保護されなければならないもうひとつのものとして，著作権がある。著作権とは，著作物を創作した著作者が有する権利のことである。著作物とは，「思想又は感情を創作的に表現したものであって，文芸，学術，美術又は音楽の範囲に属するもの」をいう。著作物が電子化されることにより，それらの加工・処理が容易となり，しかも情報通信ネットワークを通して広範に流通し合うという便利さが生じる。このことは情報化社会からの恩恵の一つである。例えば，インターネットは著作物の宝庫であり，簡単にそれらの著作物を楽しむことができる。しかもそれらの著作物が電子化されているために，簡単に利用したり加工したりすることが可能である。しかし一方で，このことは，それを創作した著作者の権利を侵すことにもなる危険をはらんでいる。著作者の権利を保護するために著作権法が制定されており，それによって著作物の利用について，厳格な制限が設けられている。

〈課題８〉インターネットから，著作権についての情報を収集し，著作権の理解を深めよう。
『例えば，http://www2.justnet.ne.jp/~junko_honkawa/japet/index.html』

２．参考サイト

・乱数について
http://www.bunkyo.ac.jp/~nemoto/lecture/simulation/97/random/
・著作権について
http://www2.justnet.ne.jp/~junko_honkawa/japet/index.html