文字列

スライス ―

インデックスの次に学習するのは、スライス (slice) です。文字列の部分を、連続あるいは 一定周期で新じい文字列として取り出すものであり、2種類の形式があります。

s[i:j] … s[i] からs[j-1] までの並び s[i:j:k] ··· s[i] からs[j-1] までのkごとの並び

スライス演算子を使ったスライス式 (slicing) は、[]の中に1個または2個のコロン:が入 る形式です。指定するのは、以下の値です。

… 取り出す範囲の先頭要素のインデックスに相当 開始i

… 取り出す範囲の末尾要素の次の要素のインデックスに相当 終了i

ステップ k … 取り出す際の刻み幅

取出しは、"s[j]まで"ではなく、"s[j]の直前の要素まで"です (Column 6-1)。

また、kが負であれば、末尾から先頭側へと逆方向に取り出されます。

それでは、実際に確かめてみます。アルファベット大文字 26 個が並んだ文字列 s からの取出 しを行います。右側の図と見比べながら、理解しましょう。

例 6-1 文字列とスライス式

```
>>> s = 'ABCDEFGHIJKLMNOPQRSTUVWXYZ'
>>> s[Ø:6] [
                                             01234567890123456789012345
'ABCDEF'
>>> s[Ø:1Ø:2] [
                                             ABCDEF
'ACEGI'
                                             ABCDEFGHIJKLMMDPORETUVEKYZ
>>> s[5:20:3] []
                                             ARCD FGHIJKLMNOPQRSTUVWKYZ
'FILOR'
                                             ABCDE GHIJKLM OPORSTUVZXYZ
>>> s[12:5:-1] []
'MLKJIHG'
```

さて、i, j, kの指定には、次の規則が適用されます。

- ■iとjは、len(s)よりも大きければ、len(s)が指定されたものとみなされる。 インデックスとは異なり、正当な範囲外の値を指定してもエラーとならない。
- ■iが省略されるか None であれば、Ø が指定されたものとみなされる。
- i が省略されるか None であれば、len(s) が指定されたものとみなされる。

複雑に感じられるでしょうが、実は、この規則のおかげで、極めて簡潔な指定が行えるよう になっています。いくつかの例で見ていきます。

全要素

全要素を取り出すスライスは、iが \emptyset で、jがlen(s)である $s[\emptyset:26]$ です。ただし、jのみ を省略した $s[\emptyset:]$ でも表せますし、さらにiも省略したs[:]でも表せます。

・ある要素から末尾まで

インデックスiの要素から末尾までを取り出すには、jを省略して、s[i:]とします。

末尾のn要素を取り出すのは、s[-n:]となります。

i, j, kの1個以上を省略するパターンのいくつかをまとめると、次のようになります。

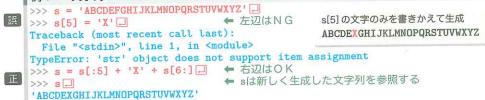
s[:]	すべて	s[:]	'ABCDEFGHIJKLMNOPQRSTUVWXYZ'
s[:n]	先頭のn要素	s[:5]	'ABCDE'
s[i:]	s[i] から末尾まで	s[5:]	'FGHIJKLMNOPQRSTUVWXYZ'
s[-n:]	末尾のπ要素	s[-5:]	'VWXYZ'
s[::k]	k - 1個おき	s[::3]	'ADGJMPSVY'
s[::-1]	すべてを逆向き	s[::-1]	'ZYXWVUTSRQPONMLKJIHGFEDCBA'

▶ nが要素数を超える場合は、全要素が取り出されます。

インデックス式とスライス式は、次章以降で学習する《リスト》や《タプル》などでも利用さ

さて、文字列はイミュータブルであって、値は変更できません。そのため、代入の左辺にイン デックス式やスライス式を置くと、エラーになります。

例 6-2 文字列 'ABCDEF…XYZ' の 'F' を 'X' に書きかえた文字列を生成



'F' を 'X' に書きかえようとして、s[5] への代入を行いますが、エラーが発生します。

'E' までのスライスと、'X' と、'F' 以降のスライスを連結させることで、うまくいきます。

Column 6-1

range とスライス式における指定値について

第4章で学習した range(n) と range(a, b) は、生成する数列の最終値が n-1 と b-1 であり、実 引数で指定した値であるnあるいはb自身は含まれません(一つ手前までとなります)。ここで学習して いるスライス式 s[:n] と s[a:b] も同様です。

このような仕様となっているのは、以下のように使利だからです。

- ■nそのもの、あるいは b a によって、長さ(要素数)が分かる。
- range(a, b) と range(b, c)、あるいは、s[:n] と s[n:] によって、途切れることなく、かつ、重 複することなく要素を列挙できる。