

理学部数学科 2000 前期 計算数学 1¹担当：辻下 徹²

今日の予定

- 文法の内容 (続き)
- クリーネの定理

参考：正規表現

UNIX では、文字列のパターンを正規表現によって表示する。これはコンピュータを用いる者には必修事項と言ってよいほど重要な実用的知識といえる。

以下 Mule というエディタについて述べる。特殊文字としては

\backslash $.$ $^$ $[$ $]$ $-$ $*$ $+$ $?$

などがある

記号	意味
$.$	任意の一文字
$\backslash m$	特殊文字 m
$[w]$	文字列 w 内の任意の一文字
r^*	正規表現 r の 0 回以上のくり返し
r^+	正規表現 r の 1 回以上のくり返し
r	正規表現 r の 0 回または 1 回のくり返し
$r1\ r2$	正規表現 $r1$ または $r2$ を選択
$[^w]$	文字列 w にはない任意の一文字
$[\alpha - \beta]$	α と β のコードの間にあるコードを持つ文字
$[a - z0 - 9]$	数と小文字の英字のいずれか
$\backslash(, \backslash)$	正規式のグループ分け

例

¹URL:<http://fcs.math.sci.hokudai.ac.jp/doc/announce/cs00.html>
質問提出アドレス:cs2000@fcs.math.sci.hokudai.ac.jp

²Email:tujisita@math.sci.hokudai.ac.jp,
Homepage:<http://fcs.math.sci.hokudai.ac.jp/tjst/>

正則表現	マッチする文字列
abc	abc
ab\\	ab\
a.b	aab, abb, a1b, aXb, a\b,...
a*b	b,ab,aab,aaab,aaaab,...
a.*b	a で始まり b で終わる任意の文字列
a[12]b	a1b,a2b のみ
a[[^] 0-9]*b	a で始まり b で終わる文字列で数字を含まないもの
x(ab cd)?	x,xab,xcd のみ
\([ab]+)\1	aa,bb,aaaa,abab,baba,bbbb,...

正則表現を用いると文字列のかなり込み入った置き換えが簡単にできる。例えば、aaawbbbvccc を aaavbbbwccc に置き換えたい場合 (w,v は aaa,bbb,ccc を含まない文字列とする) は

```
aaa\(.*\)bbb\(.*\)ccc}
```

を

```
aaa\2bbb\1ccc
```

に置き換えるとすればよい。

第 6 回の質問

[Q7-1]⁽²²⁹⁵⁰⁰³⁵⁾ パラメータ付き力学系のパラメータは t (時間) という意味だけか。

[A7-1] t 以外のパラメータである。例えば、振り子であれば、ヒモの長さ、重りの重さなどがパラメータ。

[Q7-2]⁽²²⁹⁶⁰⁰¹²⁾ 数学、形式言語と文法、離散数学とはどう関連させているか。

[A7-2] 「質問理由」が書いてないので質問の意図が不明です。どういう説明を望んでいるのかを説明して下さい。

[Q7-3]⁽²²⁹⁶⁰⁰¹⁸⁾ 書き換えは具体的にどう使われるのか？

[A7-3] 計算は書き換えの一種です。「計算」の意味を広げるもの、と考えてください。

[Q7-4]⁽²²⁹⁷⁰⁰³⁸⁾ 6.3.2 の例題の解答例 1 が $\bullet 1^n \bullet \rightarrow * 1^{2n}$ を満たすことはどう証明できるか。

[A7-4] 途中のステップを書き上げて、帰納法で示せます。

[Q7-5]⁽²²⁹⁷⁰⁰⁴¹⁾ 一回の変換で動く場所はなるべく少なくしていくようなのが良いのか？

[A7-5] そうとは限りません。なにごとにも「良し悪し」は意図に依存しています。動く場所は広げ

てもよいが、回数は減らしたい、という問題意識もあります。

[Q7-6]₍₂₂₉₈₀₀₀₃₎ $n = |\Sigma|$, $W_m := \left\{ w \in \Sigma^* \mid \ell(v) \leq m \text{ なる任意の } v \in \Sigma^* \text{ に対して } v \text{ は } w \text{ の部分語} \right\}$ としたとき、 W_m の元の長さの最小値を n, m で表せ。

[A7-6] センスのよい難問です。

[Q7-7]₍₂₂₉₈₀₀₀₇₎ Tsetlin のオートマトンについての疑問 (略)。

[A7-7] 質問の意味が不明でした。

[Q7-8]_(s980033) w における出現箇所 (occurrence) を具体的に示すアドレスのようなものはありますか？

[A7-8] 「最初から何番目」あるいは「最後から何番目」など。

[Q7-9]_(s980008) Σ^* は、 Σ によらないですよね？例えば、 $\Sigma = \{0, 1\}$ のとき、 Σ -word を $0, 1, 00, 10$ とした場合と $0, 1, 00, 10, 11$ とした場合では、 Σ^* は違いますよね。

[A7-9] Σ^* はすべての語の全体です。「 Σ -word を $0, 1, 00, 10$ とした場合」という言い方はできません。

[Q7-10]_(books) 6 . 3 . 1 「書換規則」の部分についての質問なのですが。直接書換の定義の部分で「 $W = (w_1) w (w_2)$ かつ $V = (w_1) v (w_2)$ 」とわざわざ定義したのは、何故なのでしょうか.....？

(質問理由：最初、この定義を見て、すぐにはピンと来なかったので、 w_p と言う分かりやすい書き方をせずに、こんな回りくどい言い方を何故したのが気になったので、質問にする事にしました。この場合、 w_1 は接頭語、 w_2 は接尾語の様なものですか？また、わざわざこのような書き方をしたのは、「1力所だけ v を w に書き換える」と言う事を示したかったからだ、と僕は理解したのですが、それでいいんでしょうか？)

[A7-10] 「文字列の一部に w があるときにそこだけ v に書き換える」で十分ですが、「そこだけ v に書き換える」ということを敢えて数学的に正確に書こうとすれば「 $W = (w_1) w (w_2)$ かつ $V = (w_1) v (w_2)$ 」という書き方になります。「 w_p と言う分かりやすい書き方」は色々な身ぶりの説明が残っていて数学的定義にはなっていません (もっとも、実際には身ぶりの説明を無くせるわけではないのですが)。

[Q7-11]₍₂₂₉₈₀₀₁₂₎ 1つの規則の下で、複数の安定した落ち着き先があるというのは、何か不都合を起こさないか。

[A7-11] 書き換え規則の使い方によっては不都合を起こす。後で取り上げる、「不都合」を起こす規則から不都合をなくす方法を与える。

[Q7-12]₍₂₂₉₈₀₀₁₃₎ 書き換え規則の例題 6.3.2 のような問題で用いる文字数の最小値は求められるか。

[A7-12] 難しい問題だが、パズルの。

[Q7-13]₍₂₂₉₈₀₀₁₈₎ $w_1 w_2$ から w_1 を取り除いて w_2 を作る操作はどう表すか？

[A7-13] 左から割るということで $w_1 \setminus w_1 w_2$ と書くことがある。

[Q7-14]₍₂₂₉₈₀₀₁₈₎ $\Sigma = \{ \text{英単語} \}$ のとき、接頭語は何か？

[A7-14] [The dog is old] の接頭語は [The],[The dog],[The dog is], [The dog is old] です。

[Q7-15]₍₂₂₉₈₀₀₃₀₎ 01010 の接頭語は 6 個はなぜか。

[A7-15] $\lambda, 0, 01, 010, 0101, 01010$ の 6 個です。

[Q7-16]₍₂₂₉₈₀₀₃₀₎ 01010 の 010 の出現箇所が違っていれば 2 個と数えるのか。

[A7-16] 確かに、それは部分語という言葉の使い道による、と言うべきものです。講義では、出現箇所も部分語についての情報の一部と考えます。同じ配列の DNA でも脳内にあるものと足にあるものとは別のものと考えようなものです。

[Q7-17]₍₂₂₉₈₀₀₃₄₎ 「 $(v, w) \in R$ のとき $R : v \rightarrow w$ 」がわからない。 \rightarrow_R^* とは何ですか。

[A7-17] $v \rightarrow_R w$ は $(v, w) \in R$ の別表記です。約束さえ明確にしておけば、極端には $v - R - w$ でもよいし $v/R/w$ でもよいわけです。 v を w に書き換える、というときの鍵となる情報は v, w の対だけで「●●に書き換える」という部分は、今の文脈では共通ですので、省いて良い訳です。

$v \rightarrow_R^* w$ は直接書き換えの有限列 $v \rightarrow_R v_2 \rightarrow_R \cdots \rightarrow_R w$ があることです。

[Q7-18]₍₂₂₉₈₀₀₃₅₎ $R : v \rightarrow w$ で $a \rightarrow_R^* b$ のとき $R^t : w \rightarrow v$ で $b \rightarrow_{R^t}^*$ となるか。

[A7-18] もちろんなります。

[Q7-19]₍₂₂₉₈₀₀₄₁₎ 「天皇を中心とする神の」「国」と「天皇を中心とする」「神の国」では意味が違うので、結合律が成り立たないのではないか？

[A7-19] これは語だけの問題ではなく「文法」に関連することです。品詞分解に相当する概念 parsing がありますが、それが唯一ではない、という場合が関連します。しかし、いずれにせよ、形式言語論では「意味」は考えません。(人工知能では重要なテーマですが。)

[Q7-20]₍₂₂₉₈₀₀₄₃₎ Σ^* に逆元は存在するか？

[A7-20] 空語以外の語には逆元はない。

[Q7-21]⁽²²⁹⁸⁰⁰⁴⁷⁾ 01/1, 11/0 の / の右の 0,1 は何か？

[A7-21] 入力 01 のときの出力が 1, 入力 11 のときの出力が 0, などを意味する。

[Q7-22]^(22980051,22007806) $\lambda \rightarrow v$ という書き換えを考えないのか。

[A7-22] $\lambda \rightarrow v$ による書き換えは、定義をそのまま使うと、何も無い所に v を挿入するということになります。意味はありますが、 v が限りなく増殖するだけです。通常は入れないことにしています。

[Q7-23]⁽²²⁹⁹⁰⁰¹⁹⁾ 書き換え規則のメリットは何か？

[A7-23] 規則を使うメリット以前に、書き換えという考え方が重要です。計算とは何か、という問いに対する一つの答えとなっています（計算 = 書き換え）。これを認めると「計算のメリットは何か」という問いになります。

[Q7-24]⁽²²⁹⁹⁰⁰⁴²⁾ カオスとは何か？

[A7-24] 規則と不規則の両面を持つ挙動を総称してカオスということがあります。さいころ投げの結果はカオスとは言いません。しかし、規則に従う変化の中にさいころ投げ的な振舞が現れるとき、カオス的、ということがあります。