計算の理論 I
正則表現

火曜3校時
大月 美佳

今日の講義内容

1. 正規 (正則) 表現
   1. 正則表現を使うと嬉しいこと
   2. 前準備
   3. 正則表現の定義
   4. 正則表現の例
2. ミニテスト

正則表現を使えると嬉しいこと

- パターンマッチにしょっちゅう使う
  - UNIXのシェル
    - sh, csh, ksh, tcsh, bash
    - ファイルの名前
  - UNIXのコマンド、プログラミング言語
    - egrep, awk, sed, perl
    - ファイルの中の文字列の処理

Bash のファイル名展開

% ls

aaa, aab, aac, aba, aca, ada, abc, bbb, bcb

% echo a[abc]a

aaa aba aca

Perl での処理

Sun May 27 21:51:40 JST 2001

↓

s/¥w+ (¥w+) (¥d¥d) ¥d¥d:¥d¥d:¥d¥d ¥w+ (¥d+)/Today is ¥1 ¥2, ¥3./;

↓

Today is May 27, 2001.

前準備 その1 (記号列の集合)

- アルファベット Σ
- Σ上の記号列 Σ*
- Σ*の部分集合 L, L_1, L_2

Σ: a, b, c

Σ*: ε, a, b, c, ab, ac, ba, bc, ca, cb, aab, aba, ….

L: ε, ab, bc, aba, bcca

L_1: a, bc, aca

L_2: ba, abc

前準備 その2 (記号列の集合の演算)

- $L_1L_2 = \{xy \mid x \in L_1, y \in L_2\}$ ←連接

L_1: a, bc, aca

L_2: ba, abc

⟶

$L_1 L_2$: aba, bcba, acaba, aabc, bcabc, acaabc

前準備 その3 (記号列の集合の演算)

$$\begin{cases} L^0 = \{\varepsilon\} \\ L^i = LL^{i-1} \quad (i \geq 1) \end{cases}$$

L^0: ε

$L^1 = L$: ba, abc

$L^2 = LL^1 = LL$: baba, baabc, abcba, abcabc

$L^i = LL^{i-1} = LL...L$: ba…ba, ba…abc, … …, abc…abc

前準備 その4
(Kleene閉包)

$$L^* = \bigcup_{i=0}^{\infty} L^i$$ LのKleene閉包（または単に閉包）

closure　closure

$$= L^0 \cup L^1 \cup \ldots \cup L^i \cup \ldots \cup L^\infty$$

L^0: ε ∪ L^1: ba, abc ∪…∪ L^∞: ba…ba, ba…abc, … …, abc…abc

↓

L^*: ε , ba, abc, …, ba…ba, ba…abc, …, abc…abc

前準備 その5
(正閉包)

$$L^+ = \bigcup_{i=1}^{\infty} L^i$$ Lの正閉包

positive closure

L^1: ba, abc ∪…∪ L^∞: ba…ba, ba…abc, … …, abc…abc

↓

L^+がεを含むのはLがεを含むときのみ

L^+: ba, abc, …, ba…ba, ba…abc, …, abc…abc

例

$$L_1 = \{10, 1\}$$
$$L_2 = \{011, 11\}$$
$$L_1 L_2 = \{10011, 1011, 111\}$$

$$\{10, 11\}^* = \{\varepsilon, 10, 11, 1010, 1011, 1110, 1111, \ldots\}$$
$$\{10, 11\}^+ = \{10, 11, 1010, 1011, 1110, 1111, \ldots\}$$

正則表現の定義

1. **φは正則表現で、その表す集合は空集合である。**
2. **εは正則表現で、その表す集合は{ ε }である。**
3. **Σの各元aに対してaは正則表現で、その表す集合は{a}である。**
4. **rとsがそれぞれ言語RとSを表す正則表現のとき、(r+s)、(rs)、および(r*)は正則表現で、その表す集合はそれぞれ、R∪S、RS、R*である。**

正則表現の例

- φ、ε、a
- (φ+ε)=φ∪{ε}={ε}
- (φ+a)=φ∪{a}={a}
- (ε+a)={ε}∪{a}={ε, a}
- (a+b)={a}∪{b}={a, b}
- <u>φε={}{ε}={}=φ</u>
- <u>φa={}{a}={}=φ</u>
- εa={ε}{a}={a}
- ab={a}{b}={ab}
- φ*={}*={ε}
- ε*={ε}*={ε}
- a*= {a}*={ε, a, aa, aaa, …}
- ((φ+ε)+a)=(ε+a)={ε}∪{a}={ε, a}
- <u>(φ(a+b))={}{a, b}={}=φ</u>
- ((ab)+ε)={ab}∪{ε}={ε, ab}
- ((ab)*)={ab}*={ε, ab, abab, ababab, … }

間違いやすい

正規表現の演算の強さ

* > **連接** > +

- ((0(1*))+0) → 01*+0
- (1+(10))* → (1+10)*
- ((1(1(1*)))+(01)) → 111*+01

正則表現と集合の例1

- 00={00}
- (0+1)*={0, 1}*
  ={ε, 0, 1, 00, 01, 10, 11, …}
- (1+10)*={1, 10}*
  ={ε, 1, 10, 11, 110, 101, 1010, … }
- (0+ε)(1+10)*={0, ε}{1, 10}*
  ={0, ε} {ε, 1, 10, 11, 110, 101, 1010, … }
  = {ε, 0, 1, 01, 10, 010, 11, 011, 110, 0110, 101, 0101, 1010, 01010, … }

正則表現と集合の例2

- (0+1)*011={0, 1}*011
  ={ε, 0, 1, 00, 01, 10, 11, …}{011}
  ={011, 0011, 1011, 00011, 01011, 10011, 11011, … }
- 0*1*2*={0}*{1}*{2}* 図2.8のNFA
  ={ε, 0, 00, … }{ε, 1, 11, … } {ε, 2, 22, … }
  = {ε, 0, 1, 01, 012, 00, 001, 0011, 0012, 00112, 001122, 000, 0001, 00011, 00012, 000111, 000112, 0001112, 00011122, 000111222, … }

(1+10)*の性質

- 1で始まり、連続した0を含まない列か空列から成る集合

→帰納法で示す。

$(1+10)^* = \{1,10\}^* = \bigcup_{i=0}^{\infty} \{1,10\}^i$

$i = n$のとき、$\bigcup_{i=0}^{n} \{1,10\}^i$が

1で始まり連続した0を含まない列か空列のみから成るためには

$\{1,10\}^0, \dots, \{1,10\}^n$がそれぞれ

連続した0を含まない列か空列のみからなっていればよい。

帰納法での証明つづき

1. $i = 0$のとき、$\{1,10\}^0 = \{\varepsilon\}$
2. $i = 1$のとき、$\{1,10\}^1 = \{1,10\}$

   で、1で始まり連続した0を含まない列のみを含む。
3. $i = \mathrm{n}$のとき、

$\{1,10\}^{\mathrm{n}}$が$\{1x_1, \dots, 1x_k\}$のように

1で始まり連続した0を含まない列のみを含む

とすれば($x_1, \dots, x_k$は連続した0を含まない列)、

$i = n+1$のとき、

$\{1,10\}^{n+1} = \{1,10\}\{1,10\}^n = \{11x_1, 101x_1 \dots, 11x_k, 101x_k\}$

となり、1で始まり連続した0を含まない列のみを含むようになる。

よって1,2,3より、

$(1+10)^*$は1で始まり連続した0を含まない列か空列のみを含む。

その他の演算について

$r + s = s + r$

$(r + s) + t = r + (s + t)$

$(rs)t = r(st)$

$r(s + t) = rs + rt$

$(r + s)t = rt + st$

$\phi^* = \varepsilon$

$(r^*)^* = r^*$

$(\varepsilon + r)^* = r^*$

$(r^* s^*)^* = (r + s)^*$

証明できるかな？

ミニテストと次回内容

- **ミニテスト**

  教科書・資料を見ても、友達と相談しても良い

  10分後に指名された人は板書
- **ミニテストを提出すること**

  出したら帰って良し
- **次回(6/15)内容**

  正則表現とFAとの等価性