



計算の理論 I

正則表現

火曜3校時
大月 美佳

平成16年6月8日

佐賀大学知能情報システム学科

1

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

平成16年6月8日 佐賀大学知能情報システム学科 2

正則表現を使えると嬉しいこと

- ◆ パターンマッチにしおりゅう使う
 - 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

平成16年6月8日 佐賀大学知能情報システム学科 3

Bash のファイル名展開

```
% ls
aaa , aab , aac , aba , aca , ada , abc , bbb ,
      bcd
```

```
% echo a[abc]a
aaa aba aca
```

平成16年6月8日 佐賀大学知能情報システム学科 4

Perl での処理

Sun May 27 21:51:40 JST 2001

s/Yw+ (Yw+) (YdYd) YdYd:YdYd:YdYd Yw+
(Yd+)/Today is ¥1 ¥2, ¥3./;

Today is May 27, 2001.

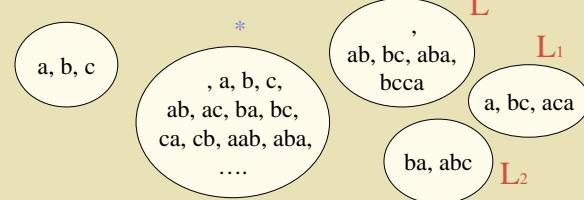
平成16年6月8日

佐賀大学知能情報システム学科

5

前準備 その1 (記号列の集合)

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



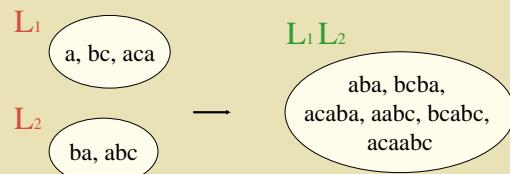
平成16年6月8日

佐賀大学知能情報システム学科

6

前準備 その2 (記号列の集合の演算)

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



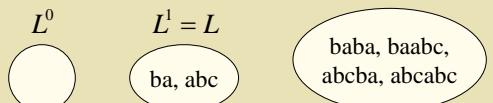
平成16年6月8日

佐賀大学知能情報システム学科

7

前準備 その3 (記号列の集合の演算)

$$\begin{cases} L^0 = \{\varepsilon\} \\ L^i = LL^{i-1} \quad (i \geq 1) \end{cases} \quad L^2 = LL^1 = LL$$



$$L^i = LL^{i-1} = LL...L$$

ba...ba, ba...abc, ...
..., abc...abc

平成16年6月8日

佐賀大学知能情報システム学科

8

前準備 その4 (Kleene閉包)

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

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

$$\begin{array}{ccccccc} L^0 & & L^1 & & & & L^\infty \\ \text{---} & & \text{---} & & & & \text{---} \\ \text{ba, abc} & & \dots & & \text{ba...ba, ba...abc, ...} & & \dots, abc...abc \\ | & & & & & & | \\ L^* & & & & & & L^+ \end{array}$$

, ba, abc, ..., ba...ba, ba...abc, ..., abc...abc

平成16年6月8日

9

前準備 その5 (正閉包)

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

$$\begin{array}{ccccccc} L^1 & & & & & & L^\infty \\ \text{---} & & & & & & \text{---} \\ \text{ba, abc} & & \dots & & \text{ba...ba, ba...abc, ...} & & \dots, abc...abc \\ | & & & & & & | \\ L^+ & & & & & & L^+ \end{array}$$

ba, abc, ..., ba...ba, ba...abc, ..., abc...abc

平成16年6月8日

10

例

$$L_1 = \{10, 1\}$$

$$L_2 = \{011, 11\}$$

$$L_1 L_2 = \{10011, 1011, 111\}$$

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

$$\{10, 11\}^+ = \{10, 11, 1010, 1011, 1110, 1111, \dots\}$$

平成16年6月8日

佐賀大学知能情報システム学科

11

正則表現の定義

1. ε は正則表現で、その表す集合は空集合である。
2. a は正則表現で、その表す集合は $\{a\}$ である。
3. a 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 $R \cup S$ 、 RS 、 R^* である。

平成16年6月8日

佐賀大学知能情報システム学科

12

正則表現の例

間違いやすい

平成16年6月8日

佐賀大学知能情報システム学科

13

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

平成16年6月8日

佐賀大学知能情報システム学科

14

正則表現と集合の例1

- ◆ $00 = \{00\}$
 - ◆ $(0+1)^* = \{0, 1\}^*$
 $= \{ \quad , 0, 1, 00, 01, 10, 11, \dots \}$
 - ◆ $(1+10)^* = \{1, 10\}^*$
 $= \{ \quad , 1, 10, 11, 110, 101, 1010, \dots \}$
 - ◆ $(0+1)(1+10)^* = \{0, \quad \} \{1, 10\}^*$
 $= \{0, \quad \} \{ \quad , 1, 10, 11, 110, 101, 1010, \dots \}$
 $= \{ \quad , 0, 1, 01, 10, 010, 11, 011, 110, 0110, 101,$
 $0101, 1010, 01010, \dots \}$

平成16年6月8日

佐賀大学知能情報システム学科

15

正則表現と集合の例2

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

平成16年6月8日

佐賀大学知能情報システム学科

16

(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。

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

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

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。

平成16年6月8日

佐賀大学知能情報システム学科

17

帰納法での証明つづき

- $i = 0$ のとき、 $\{1,10\}^0 = \{ \}$
- $i = 1$ 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。
- $i = n$ のとき、
 $\{1,10\}^n$ が $\{x_1, \dots, x_k\}$ 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。

平成16年6月8日

佐賀大学知能情報システム学科

18

その他の演算について

$$r + s = s + r$$

$$* = \varepsilon$$

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

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

$$(rs)t = r(st)$$

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

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

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

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

証明できるかな？

平成16年6月8日

佐賀大学知能情報システム学科

19

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

平成16年6月8日

佐賀大学知能情報システム学科

20