

正規表現のフォーム記述言語への応用

An Approach to Design Form Definition Languages

広島大学 工学部 杉原 一夫 (Kazuo SUGIHARA)
 宮尾 淳一 (Jun'ichi MIYAO)
 高山 雅行 (Masayuki TAKAYAMA)
 菊野 亨 (Tohru KIKUNO)

1. まえがき

オフィスオートメーションを目指した情報管理システムにおける重要な課題の1つにマンマシンインタフェースの開発がある。特に文書作成におけるフォーマットの柔軟な記述に関心が集まっている⁽¹⁾⁽³⁾。これに関連して、フォーム(Form)をユーザインタフェースの中心的媒体として使用する試みがなされ、例えば OBE, Officetalk などが開発されてきている⁽⁴⁾。最近では、ユーザの要求する多様なフォームをスクリーン上に自由に表示する試みも行われている⁽⁵⁾。

本稿では、標準レポート、定形書簡および文書作成に使用することを目的として設計したフォーム記述言語について述べる。先ず、2. で基礎的な考察としてフォーム、及び、フォーム上での変換を定義する。次に、3. において、フォーム上の各変換を具体的に与えるためのフォーム記述言語を提

案する。この言語による記述はregion部とskeleton部の2つの階層記述からなる。region部の記述には正規表現が用いられる。一方、skeleton部ではテンプレートを用いて、フォームを2次元構造で容易に指定できる。最後に4.では、記述例を用いてフォームの具体的な説明を行う。

なお、文献(3)において、本稿と同様の試みが行われているが、それとの関連については5.で簡単に触れる。

2. フォーム

ここでは、アルファベットを Σ で表し、属性を $A_1, A_2, \dots, A_n$ で表す。このとき、属性の有限集合をスキーム $S = \{A_1, A_2, \dots, A_n\}$ (各 $A_i \in \Sigma$ とする)と定義する。各 A_i は Σ^* の要素を値として持つものとし、 A_i の取り得る値の集合を $D(A_i)$ で表し、 A_i の定義域と呼ぶ。

[定義1] スキーム $S = \{A_1, A_2, \dots, A_n\}$ と定義域 $D(A_i)$ に対し、2項組の集合 $F_0 = \{(A_i, \alpha_i) \mid \alpha_i \in D(A_i), 1 \leq i \leq n\}$ をスキーム S の1つの抽象フォームと呼ぶ。

2項組の集合 $I = \{(w_i, \alpha_i) \mid 1 \leq i \leq m\}$ を割り付けと呼ぶ。ここで各 w_i はウィンドウであり、 α_i はアルファベッ

$\Sigma' = \Sigma \cup \{A_1, A_2, \dots, A_n\}$ 上の文字列である。

[定義 2] 抽象フォーム $F_0 = \{(A_i, \alpha_i) \mid \alpha_i \in D(A_i), 1 \leq i \leq n\}$ と割り付け $I = \{(w_j, a_j) \mid a_j \in (\Sigma')^*, 1 \leq j \leq m\}$ に対し, 次の条件を満たす 2 項組 $F_1 = \{(w_j, a'_j) \mid a'_j \in \Sigma^*, 1 \leq j \leq m\}$ を F_0 の I に関する中間フォームと呼ぶ。

条件: 系列 a'_j は系列 a_j 中に現われる各文字 A_i を対応する系列 α_i ($(A_i, \alpha_i) \in F_0$) で置き換えて得られる文字列である[†]。

スクリーン $\mathcal{S}$ を集合 $\{(i, j) \mid 1 \leq i \leq x, 1 \leq j \leq y\}$ で表す。ここで x はスクリーンの行数で y は列数である。各 (i, j) をピクセルと呼ぶ。

スクリーン $\mathcal{S}$ と F_0 と I に関する中間フォーム $F_1 = \{(w_i, a'_i) \mid 1 \leq i \leq m\}$ に対して, 次の条件を満たす $\mathcal{S}' = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \dots \cup \mathcal{S}_m$ を I の $\mathcal{S}$ 上でのレイアウトと呼ぶ。

条件: ① $\mathcal{S}_i \subseteq \mathcal{S}$ ($1 \leq i \leq m$)。

② 各 $\mathcal{S}_i$ は $\mathcal{S}$ 上の複合長方形領域 (垂直, 水平の線分

[†] 対応する系列を用いた, 系列 a_i から系列 a'_i への変換の詳細な説明は 3.1 で与える。

で囲まれた閉領域)であり, しかも $\mathcal{X}_i$ 中のピクセル数と文字列 A_i の長さが一致する.

レイアウト $\mathcal{X}$ 中の $\mathcal{X}_i$ をウィンドウ w_i に対する部分スクリーンと呼ぶ.

[定義3] F_0 の I に関する中間フォーム $F_1 = \{ (w_i, A_i) \mid 1 \leq i \leq m \}$ と I の $\mathcal{X}$ 上でのレイアウト $\mathcal{X}' = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \dots \cup \mathcal{X}_m$ に対し, 次の条件を満たす項組 $F_2 = \bigcup_{i=1}^m \{ (j, t, a_e) \mid (j, t) \in \mathcal{X}_i, a_e \in \Sigma \}$ を F_1 の $\mathcal{X}'$ に関する実在フォームと呼ぶ.

条件: $\mathcal{X}_i$ の要素 (j_1, t_1) と (j_2, t_2) に対し, 大小関係を $j_1 < j_2$ あるいは $j_1 = j_2$ かつ $t_1 < t_2$ のとき, $(j_1, t_1) < (j_2, t_2)$ と定める. このとき, $\mathcal{X}_i$ の x 番目に小さい要素を (j, t) とし $(j, t, a_e) \in F_2$ とすると, a_e は A_i の左から x 番目の文字である.

抽象フォームと中間フォームがいずれも論理的な, あるいは, 一次元的なデータ構造を取り扱っているのに対し, 実在フォームではスクリーン(ディスプレイ装置の表示画面)上の物理的なデータ構造を決めている. 通常, フォームと呼ばれているのは実在フォームのことである.

3. フォーム記述言語

フォーム F_2 を求めるには、先ず、与えられた抽象フォーム F_0 と割り付け I を用いて中間フォームを計算する。引き続き、レイアウト $\mathcal{G}'$ の決定を行った後、最終的に F_2 を構成する。

提案するフォーム記述言語では割り付け I 中の系列 a_i の指定を正規表現を用いて行う。この部分を region 部と呼ぶ。次にレイアウト $\mathcal{G}'$ の決定は、厳密には、相対的な位置決定と物理的な埋め込みに分けて実行される。この部分を skeleton 部と呼ぶ。(相対的な位置決定から物理的な埋め込みへの変換は、埋め込まれる文字列の長さに応じて動的に行われる。)

以下、region 部と skeleton 部について順次説明する。

3.1 region 部

割り付け $I = \{ (w_i, a_i) \mid 1 \leq i \leq m \}$ を指定する部分である。系列 a_i はアルファベット $\Sigma' = \Sigma \cup \{ A_1, A_2, \dots, A_n \}$ 上の正規表現として表わされる。 Σ' 上の正規表現は次のように定義される⁽²⁾。

- (1) ϕ は正規表現であり、空集合を表す。
- (2) λ (空系列) は正規表現であり、集合 $\{ \lambda \}$ を表す。
- (3) 各要素 $a \in \Sigma'$ に対し、 a は正規表現であり、集合 $\{ a \}$ を表す。
- (4) x, y が集合 X, Y を表す正規表現であるなら、 $x|y$

と y , x^* は正規表現であり, それぞれ $X \cup Y$, XY , X^* を表す.

提案するフォーム記述言語では, 便宜上, 各 $(w_i, \alpha_i) \in I$ は $i := \alpha_i$ の形式で, また正規表現の範囲を $(,)$ で表す.

上で説明した割り付け $I = \{ (w_i, \alpha_i) \}$ に対し, 抽象フォーム $F_0 = \{ (A_i, \alpha_i) \}$ が与えられると, 次のようにして中間フォーム $F_1 = \{ (w_i, \alpha'_i) \}$ が構成される.

手順 1 (α_i の解釈): 正規表現 α_i を対応する F_0 の要素に基づいて, 連接のみを用いた表現形式 $\overline{\alpha_i}$ に変換する. 先ず, $\alpha_i = A \mid B$ に対しては, F_0 中に A , B のいずれかの属性に対して 2 項組が存在することが保証されていると仮定する. 今それを A (又は B) とすると $\overline{\alpha_i} = A$ (又は $\overline{\alpha_i} = B$) の形に変換される. 次に, $\alpha_i = A^*$ に対しては, F_0 中に一般に複数個 (n_A) の 2 項組 (A, α_1) , (A, α_2) , $\dots$, (A, α_{n_A}) が存在したと仮定する. このとき $\overline{\alpha_i} = \overbrace{A A \dots A}^{n_A}$ の形に変換する.

手順 2 (α'_i の構成): 手順 1 で求めた $\overline{\alpha_i}$ 中の各 A_i を対応する要素 α_i ($(A_i, \alpha_i) \in F_0$) で置き換える. 一般に, 同じ A_i に対し複数個 (n_i) の 2 項組 (A_i, α_{i_1}) , (A_i, α_{i_2}) , $\dots$, $(A_i, \alpha_{i_{n_i}})$ が F_0 中に存在する. 先ず, $n_i = 1$ のときは, 単に, $\overline{\alpha_i}$ 中に 1 度だけ現われる A_i を α_{i_1} で置き換える. 次に, $n_i \neq 1$ のときは, $\overline{\alpha_i}$ 中の j ($1 \leq j \leq n_i$) 番目に現われる A_i を α_{i_j} で

置き換えて行く。こうして求まった系列を A_i' と定める。

3.2 skeleton部

スクリーン $\mathcal{S}$ と中間フォーム $F_1 = \{ (w_i, A_i') \mid 1 \leq i \leq m \}$ に対し、レイアウト $\mathcal{S}' = \bigcup_{i=1}^m \mathcal{S}_i$ を指定する部分である。なお、各部分スクリーン $\mathcal{S}_i$ は $\mathcal{S}$ 上での位置と大きさをテンプレートを用いて2次元的に指定する。

テンプレートによる $\mathcal{S}_i$ の記述は、識別をID番号と下線で行い、その位置を記号@と下線による相対的な位置で指定し、その大きさは下線の長さを用いて行う。また、この $\mathcal{S}_i$ には、固定、可変、定数形式の3種類がある。先ず、固定形式の場合には、下線 $\underline{i}$ を用い下線の位置と長さで $\mathcal{S}_i$ を記述する。次に可変形式の場合には、記号@と下線を用い $\underline{i} @$, $\underline{i} @$, $\underline{i} @$ などで表す。ここで記号@は $\mathcal{S}_i$ の大きさが可変であることを表す。最後に定数形式の場合には、 Σ 上の文字列を用いて直接表す。

なお、各部分スクリーンの間の位置関係は相対的なものであり、可変形式の部分スクリーンを用いても、各部分スクリーン相互の位置関係は変化しない。

また、例外的に、ユーザからの入力を要求する部分スクリーンであることを指定することもできる。その場合には、例えば $\underline{\text{input } 1}$ のように"input"で指定する。

4. 記述例

ここでは、アメリカ特許公報の書式（図1）を例にとって、提案するフォーム記述言語の記述例を簡単に説明する。

先ず、スキーム $S = \{ \text{No}, \text{title}, \text{name1}, \text{name2}, \text{name3}, \text{street}, \text{city}, \text{state}, \text{zip}, \text{date}, \text{serialno}, \text{text} \}$ とする。

United State Patent Office	3310775
	Patented Mar.21,1980
3310775	
VEHICLE BURGLAR ALARM	
James E. Birthe, and John D. Grollman,	
1428 W. 47th St., Los Angeles, Calif. 90062	
Filed Feb.24,1979, Ser. No.434900	
ABSTRACT OF THE DISCLOSURE	
A vehicle burglar alarm having two electrical circuits, one a setting circuit having a master key switch for placing the setting circuit in condition for operation, and an alarm circuit which is brought into operation as a result of tampering with the vehicle and which will operate for a period of time independent of the setting circuit.	
N. C. READ	
A. H. WARING	

図1. アメリカ特許公報

4.1 region部

割り付け $I = \{ (w_i, a_i) \mid 1 \leq i \leq 8 \}$ の記述例を図2に示す。各ウィンドウ w_i は1～8のID番号で示す。 a_i は各ウィンドウに対する1次元構造を正規表現を用いて指定する。図2で、`'`で囲まれた文字列は直接スクリーン上に表示されることを表す。

```

1:=No
2:=title
3:=(name1',')|(name1','(' 'name1','')*'and' name1',')
4:=street','city','state','zip
5:='Filed'date', Ser. No.'serialno
6:=text
7:=name2
8:=name3

```

図2 region部

4.2 skeleton部

テンプレートを用いたレイアウト $\mathcal{S}' = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \dots \cup \mathcal{S}_8$ の記述例を図3に示す。

まず、固定形式の部分スクリーンとしては $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_5, \mathcal{S}_7, \mathcal{S}_8$ がある。これは中間フォーム $F_1 = \{ (w_i, a_i) \mid i = 1, 2, 5, 7, 8 \}$ に対し、 a_i の長さ (a_i 中に含まれる文字数) がある程度予測できる場合に用いる。

次に、可変形式の部分スクリーンとしては x_3, x_4, x_6 がある。これは $F_1 = \{ (w_i, A'_i) \mid i = 3, 4, 6 \}$ に対し、 A'_i の長さの予測が困難であると考えられる場合に用いる。例えば、 A'_4 は住所であり長さの予測ができない。そこで、記号@を用いて4_____@と指定することにより、1行の範囲内で長さが自由に変わること示す。また、 A'_6 は特許本文に対応する部分であり、その内容に応じて行数が変化する。そこで、 $\frac{6}{6}$ _____@と指定することによって行数が自由に変わること示す。同時に6_____@と指定することによって列数も可変であることを示す。更に、@_____と指定することによって、2行目以降が右端をそろえることが指定される。そして最終的には、図3の点線で囲まれた複合長方形領域が対応する。

最後に、定数形式の部分スクリーンには、United State Patent Office, Patented, ABSTRACT OF THE DISCLOSUREがある。

United State Patent Office		1 _____
		Patented <u>input1</u>
1	_____	
2	_____	
3	_____	@
4	_____	@
5	_____	
ABSTRACT OF THE DISCLOSURE		
6	_____	@
6	_____	@
@		
6	_____	@
7	_____	
8	_____	

図3 skeleton部

5. あとがき

本稿では，フォームを階層的に定義し，フォーム間で行われる変換について説明した．更に，正規表現およびテンプレートを用いて，フォームの有する2次元構造を記述するフォーム記述言語を提案した．

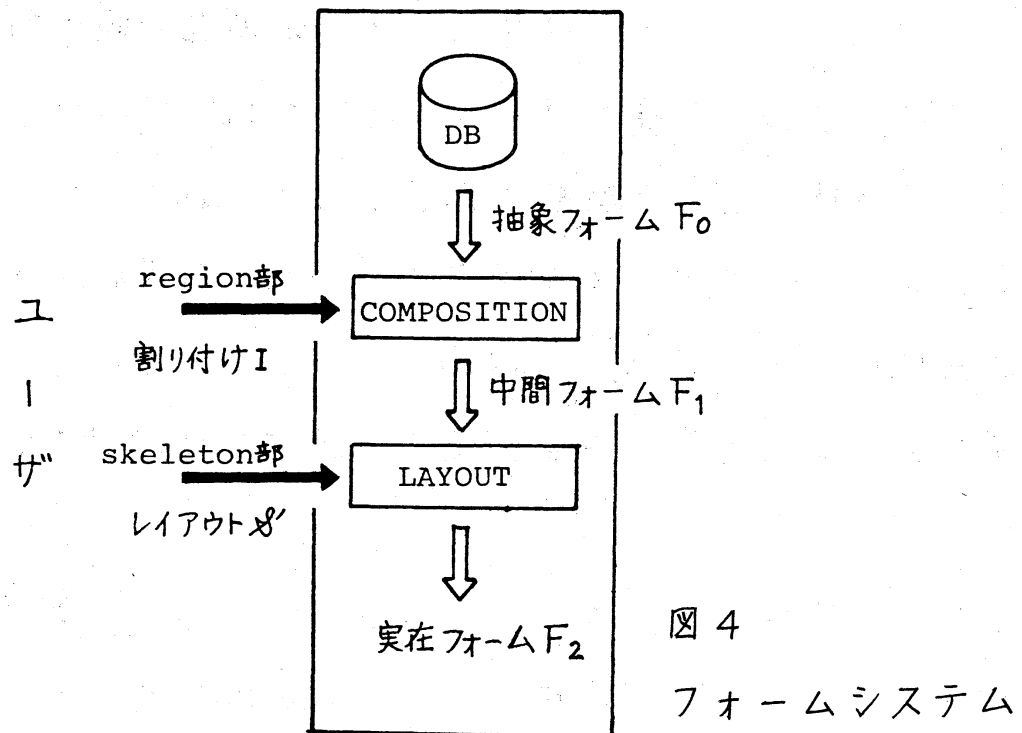
フォーム作成に関する同様の研究として文献(3)がある．文献(3)においても，正規表現を用いて抽象フォームの構造を指

定している。しかし、その構造としては木構造表現を許している。これに対し、本稿で提案した方式では、抽象フォームは属性とその持つ値の2項組であると定められている。したがって、関係データベースを仮定するなら、各属性との対応付けは直接行える。また、文献(3)において、中間フォームから実在フォームへの変換およびその記述方法は議論されていない。

図4にこの言語に基づくフォームシステムの概念図を示す。すなわち、ユーザは正規表現を用いたregion部の記述と、テンプレートを用いたskeleton部の記述を与えることにより、容易にフォームの作成ができる。

現在、図4中のモジュールCOMPOSITIONの機能の強化について検討中である。今後、フォーム記述言語の開発を行う予定である。

最後に、日頃御指導いただく本学吉田典可教授に深く感謝いたします。



文献

- (1) Furuta, R., Scofield, J. and Shaw, A.: "Document formatting systems: Survey, concepts, and issues," Computing Surveys, 14, 3, pp.417-472(1982).
- (2) Hopcroft, J.E. and Ullman, J.D.: "Introduction to Automata Theory, Languages, and Computation," Addison-Wesley(1979).
- (3) 上林弥彦: "フォームの設計問題", 信学技報, AL83-44, PP.113-121(1983).
- (4) King, K.J. and Maryanski, F.J.: "Information management trends in office automation," Proceedings of the IEEE, 71, 4, pp.519-528(1983).
- (5) Rowe, L.A. and Shoens, K.A.: "Programming language concepts for screen definition," IEEE Trans. Software Engineering, SE-9, 1, pp.31-39(1983).