

多国語自動組版における文字調整のオートマトン方式

翟 建一 (ザイ ケンイチ) †

†株式会社ロココ 〒105-0014 東京都港区芝2-2-1 4一星芝ビルディング6F

E-mail: †k_zai@rococo.co.jp

あらまし 昨今、出版印刷業界では自動組版技術が進化しつつある。多国語の自動組版の場合、翻訳で文章の文字数が変わるので、複数言語で同一文章の印刷（多国語対応）が求められる場合、印刷フォーマットを変えずに、更に人工操作も介さずに、自動的に文字調整できることが不可欠となる。

本論文は、多国語対応の自動組版の文字調整について、オートマトン (Automaton) モデルを抽出し、それを用いた、より知能的な文字調整方式の実現について論じるものである。

キーワード オートマトン、アルゴリズム、自動組版、多国語

An Automaton Method of Character Adjustment in Multi National Language Text Edition

Kenichi ZAI†

† Rococo Corporation Ichiboshi-Shiba Building 6F 2-2-14 Shiba, Minato-ku, Tokyo, 105-0014 Japan

E-mail: †k_zai@rococo.co.jp

Abstract The technology of auto edition in the publishing industry is being improved obviously now. When the same article that is written by different national language is edited, because the article is translated the number of words and characters is changed. In order to print same article that is written by different national language without any changing the format of the article, without any editing the article by human operator, then an automatic method that can adjust the character in the article is very necessary.

In this paper, about adjusting character in the article that written by translation language with automatic edition method, how to create an automaton model and how to release this intellectual method that is created by using the automaton model to adjust the character of the article has been discussed.

Keyword Automaton, algorithm, Edition, Multi National Language

1. はじめに

出版印刷業界の技術進歩に伴い、自動組版技術が要求されている。自動組版を導入する主な目的は手動組版で行っている作業をプログラムで自動化して、人工操作をせずに効率よく組版を行うとのことである。

自動組版とは、本来手動によって行われる組版作業手順をパソコン上で人工操作を介さずに自動的に文字や画像等の配置や調整などを連続で行ってくれる仕組みのことである。出版印刷業界には、自動組版の解釈は色々であるが、本論文の自動組版概念は、テンプレートファイルによって、人工操作をせずに、自動的に印刷レイアウトを組み合わせる技術である。

多国語の自動組版の場合、翻訳により、元文章と比

べ、訳文の文字数が多くなったり少なくなったりするが、複数言語で同一文章の印刷（多国語対応）が求められると、自動組版を行なう際に自動的に文字調整が不可欠となる。

印刷のフォーマットの変更や人工操作をせずに多国語対応の自動組版のソフトを開発する為に、オートマトン (Automaton) 方式を利用し、多国語対応の自動組版に文字調整機能が必要になってくる。

本論文は、文字調整のルールに則ったオートマトンモデルを定義し、オートマトン方式を利用し、より知能のある自動組版の文字調整方式を論じる。その中で特に文字溢れの調整を中心に論じる。

2. 文字調整について

2.1. 多国語対応の自動組版

ここでいう多国語対応の自動組版とは、アドビ社製品の InDesign CS に基づいて InDesign の SKD (アドビ社の開発プラットフォーム) で開発されたプラグイン及びその他ソフトからなるアプリケーションシステムのことである。

InDesign では、文字情報はテキストフレームに収納されている。翻訳 (例えば日本語→英語) によって文字が増えると、テキストフレーム内にオーバーフローが発生し、文字を表示し切れなくなってしまう。この場合、原文のフォーマット及びスタイルを維持するまま、いかに文字を調整するのが課題となる。

基本的には、発生したオーバーフローの情報と調整可能な項目をまとめて考えて、有限オートマトンのモデルを抽出して、最適な方法を求めたい。

2.2. 文字溢れの情報及び調整可能な項目

InDesign のテキストフレームでは、オーバーフローの場合、その状態だけではなく、溢れている文字数も取得できる。また一般的なテキストフレームでは、表示スタイル毎に文字比率 (X, Y)、文字サイズ、文字詰め率、字間、行間などの項目で、テキストフレームを調整できる。論述を簡潔化する為に、調整項目を3つに限定する。

有限オートマトンのモデルを抽出する為に、下記のように定義する。

定義1: テキストフレーム内の文字数は N_t とする。

定義2: テキストフレームから溢れている文字数は N_f とする。

定義3: 文字の溢れ率は R_f とする。 $R_f = N_f \div N_t$ (溢れ率 = 溢れている文字数 ÷ テキストフレームにある文字数)。

定義4: 文字サイズを調整する変動率は R_s とする。

定義5: 文字比率を調整する変動率は R_r とする。

定義6: 文字詰め率を調整する変動率は R_p とする。

R_f 、 R_s 、 R_r 、 R_p の単位は、全てパーセントとする。

定義7: R_s の変動範囲は $100\% \sim R_{sm}\%$ ($R_{sm} < 100\%$) とする。 R_{sm} とは、容認された文字サイズの調整限度である。

定義8: R_r の変動範囲は $100\% \sim R_{rm}\%$ ($R_{rm} < 100\%$) とする。 R_{rm} とは、容認された文字比率の調整限度である。

定義9: R_p の変動範囲は $0\% \sim R_{pm}\%$ ($R_{pm} < 100\%$) とする。 R_{pm} とは、容認された文字詰め率の調整限度である。

R_{sm} 、 R_{rm} 、 R_{pm} の単位は、全てパーセントとし、実際に経験により、決めてもかまいません。

2.3. 色々な調整方式

筆者はオートマトン方式を考案する前に、色々な調整方案を試したが、何れも満足できない欠点がある。

2.3.1. ツリー構造の調整方式

調整可能な諸項目から、調整ツリーを構造することができる。それにより、一般ツリーのアルゴリズムを利用して、テキストフレームを調整することができる。

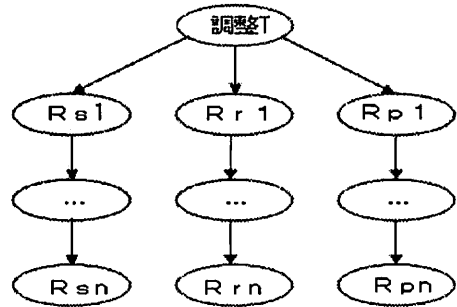


図1: 調整用ツリー

深さ優先方式を利用する場合、次のアルゴリズムとなる。

- ① ある調整項目の現在値から刻み単位で加減しながら、溢れがなくなるまで或いは限度値まで調整する。
- ② 一旦納めることができたなら直ちに調整を中止する。
- ③ ある調整項目で調整完了しても、収まらない場合、次の調整項目を取得して、調整を続ける。
- ④ 全ての調整項目がなくなるまで上記調整を続ける。何れでも調整できない場合は処理を中止する。

この場合、調整の序列は $R_{s1}, R_{s2}, \dots, R_{r1}, R_{r2}, \dots, R_{p1}, R_{p2}, \dots$ ような形となる。

幅優先方式を利用する場合、次のアルゴリズムとなる。

- ① 順番になった調整項目の現在値を取得して、刻み単位で加減し、調整項目の限度値から、超えないように、一回調整する。
- ② 一旦納めることができたなら直ちに調整を中止する。
- ③ 一回調整しても収まらない場合、順番に次の調整項目の現在値を取得して、調整を続ける。
- ④ 全ての調整項目が限度値になるまで上記調整を続ける。何れでも調整できない場合は処理を中止する。

この場合、調整の序列は $R_{s1}, R_{r1}, R_{p1}, R_{s2}, R_{r2}, R_{p2}, \dots$ ような形となる。

上記の方法は、分かりやすく、簡単に調整できる

反面、なかなか良い調整結果が得られないケースが多い。

2.3.2. 自由序列の調整方式

自由序列とは、印刷業者の経験により調整序列を指定し、実際に調整する時に指定された序列から調整項目を読み込みながら調整値を計算し、テキストフレームを調整する方式である。この方式はあるテンプレートのファイルに対してうまくいっても、全てのテンプレートファイルに適用できるとは言えない。

3. 有限オートマトン方式

前述の問題点から、多国語対応の自動組版を行う際に文字が溢れた場合、複数の属性により収納の調整処理が行われるが、どのように組み合わせるかを考えなければならない。より柔軟性のある調整序列を求めたい。より知能のあるアルゴリズムを使いたい。そこで、有限オートマトンを利用するのが有効だと考える。有限オートマトンモデルの抽出が問題を解決するポイントとなる。

本節では有限オートマトンモデルの抽出方法を論じる。

3.1. オートマトンの基礎

オートマトンとは、形式言語を処理するアルゴリズムである。即ち、コンピュータそのものを数学的観点からモデル化し、問題解決の為の処理手順を定式化したものである。具体的には入力に対して内部状態に応じた処理を行ない、結果を出力する仮想的な自動機械である。それは複数の状態及び個々の状態での入力結果に対してどのような処理を行うかを定めた関数で構成されている。具体的な動作によって、有限オートマトン、セルオートマトン、チューリングマシン、プッシュダウンオートマトンなど様々な種類があるが、本論では有限オートマトンを利用したいので、有限オートマトンの概念を記述する。

一般的に、決定性有限オートマトン(Deterministic finite automaton DFA)は、次の5項目の組で定義される。

$$M = (Q, \Sigma, \delta, Q_0, F)$$

Q: 状態の有限集合

Σ : 有限のアルファベット(入力記号の有限集合)

δ : 遷移関数 $\delta: Q \times \Sigma \rightarrow Q$

Q₀: 初期状態 $Q_0 \in Q$

F: 最終状態(あるいは受理状態)の集合 $F \subseteq Q$

有限オートマトンの論理を利用し、多国語対応の自動組版で文字を調整する為に、調整処理の流れを有限オートマトンモデル化にしなければならない。つまり自動組版システムが利用しているツールの InDesign を分析し、有限オートマトンのモデルを抽出しなければ

ならない。

3.2. 文字溢れ状態と調整ルール

従来の有限オートマトンは、入力文字によって状態遷移を決めるが、ここでは変動率と溢れ率しか利用できないので、文字の溢れ率から、文字溢れ状態を決める。但し、率というのは、連続変動するので、閾値(いきち)を設けることで、等級を分け、文字溢れ状態を定義する。

表1: 文字溢れ状態定義表

文字溢れ状態	文字溢れ率
Q5	40~100%
Q4	30~40%
Q3	20~30%
Q2	10~20%
Q1	0~10%
Qs	任意
Qe	文字溢れなし

(注: 表1の中、閾値は業務経験により決定、仮に文字溢れ状態は5段階を設定)

実際に動作する時に、一回調整した結果で新しい溢れ率を算出する。新たに算出した溢れ率によって、新しい文字溢れ状態及び次に調整する動作(オートマトンの入力記号の有限集合に相当)を決める。

次に、文字溢れ状態定義表に基づいて調整ルールを決める。(但し、閾値は一般的に業務経験に基づいて決めるので、実際の利用状況により上記の閾値及び下記のルールを変えても良い。)

- 調整能力の降順は文字サイズ→文字比率→文字詰め率とする(Rule-1)。
- ある調整項目で調整が進まない場合(調整制限を越える場合)、他の調整項目で調整を続ける(Rule-2)。
- 文字溢れ状態がQ4~Q5の場合(溢れ率が高レベル)、文字サイズ→文字比率→文字詰め率の順番で調整する(Rule-3)。
- 文字溢れ状態がQ2~Q3の場合(溢れ率が中レベル)、文字比率→文字詰め率→文字サイズの順番で調整する(Rule-4)。
- 文字溢れ状態がQ1の場合(溢れ率が低レベル)、文字詰め率→文字サイズ→文字比率の順番で調整する(Rule-5)。
- 何れも調整が進めない場合(収納処理失敗)、オートマトンの受理不能とみなし、終了状態へ遷移する(Rule-6)。

3.3. 状態遷移図

3.3.1. 動作定義

有限オートマトンのモデルを抽出する為に、調整処理(オートマトンの入力記号の有限集合に相当)を定義する。

定義10：文字サイズの変動率を刻み単位で計算し、一回テキストフレームを調整することはPsとする。

定義11：文字比率の変動率を刻み単位で計算し、一回テキストフレームを調整することはPrとする。

定義12：文字詰め率の変動率を刻み単位で計算し、一回テキストフレームを調整することはPpとする。

3.3.2. 状態遷移図

有限オートマトンは、状態遷移図や状態遷移表によって表わすことができる。

前述した文字溢れ状態定義表及び定義10～12により、下記状態遷移表と遷移関数を定義する。

表2：状態遷移表と遷移関数

溢れ状態	調整処理 Ps	調整処理 Pr	調整処理 Pp
Qs 初期状態	$\delta(Qs, Ps)=Q1$	$\delta(Qs, Pr)=Q1$	$\delta(Qs, Pp)=Q1$
	$\delta(Qs, Ps)=Q2$	$\delta(Qs, Pr)=Q2$	$\delta(Qs, Pp)=Q2$
	$\delta(Qs, Ps)=Q3$	$\delta(Qs, Pr)=Q3$	$\delta(Qs, Pp)=Q3$
	$\delta(Qs, Ps)=Q4$	$\delta(Qs, Pr)=Q4$	$\delta(Qs, Pp)=Q4$
	$\delta(Qs, Ps)=Q5$	$\delta(Qs, Pr)=Q5$	$\delta(Qs, Pp)=Q5$
	$\delta(Qs, Ps)=Qe$	$\delta(Qs, Pr)=Qe$	$\delta(Qs, Pp)=Qe$
Q5	$\delta(Q5, Ps)=Q1$	$\delta(Q5, Pr)=Q1$	$\delta(Q5, Pp)=Q1$
	$\delta(Q5, Ps)=Q2$	$\delta(Q5, Pr)=Q2$	$\delta(Q5, Pp)=Q2$
	$\delta(Q5, Ps)=Q3$	$\delta(Q5, Pr)=Q3$	$\delta(Q5, Pp)=Q3$
	$\delta(Q5, Ps)=Q4$	$\delta(Q5, Pr)=Q4$	$\delta(Q5, Pp)=Q4$
	$\delta(Q5, Ps)=Q5$	$\delta(Q5, Pr)=Q5$	$\delta(Q5, Pp)=Q5$
	$\delta(Q5, Ps)=Qe$	$\delta(Q5, Pr)=Qe$	$\delta(Q5, Pp)=Qe$
Q4	$\delta(Q4, Ps)=Q1$	$\delta(Q4, Pr)=Q1$	$\delta(Q4, Pp)=Q1$
	$\delta(Q4, Ps)=Q2$	$\delta(Q4, Pr)=Q2$	$\delta(Q4, Pp)=Q2$
	$\delta(Q4, Ps)=Q3$	$\delta(Q4, Pr)=Q3$	$\delta(Q4, Pp)=Q3$
	$\delta(Q4, Ps)=Q4$	$\delta(Q4, Pr)=Q4$	$\delta(Q4, Pp)=Q4$
	$\delta(Q4, Ps)=Qe$	$\delta(Q4, Pr)=Qe$	$\delta(Q4, Pp)=Qe$
Q3	$\delta(Q3, Ps)=Q1$	$\delta(Q3, Pr)=Q1$	$\delta(Q3, Pp)=Q1$
	$\delta(Q3, Ps)=Q2$	$\delta(Q3, Pr)=Q2$	$\delta(Q3, Pp)=Q2$
	$\delta(Q3, Ps)=Q3$	$\delta(Q3, Pr)=Q3$	$\delta(Q3, Pp)=Q2$
	$\delta(Q3, Ps)=Qe$	$\delta(Q3, Pr)=Qe$	$\delta(Q3, Pp)=Qe$
Q2	$\delta(Q2, Ps)=Q1$	$\delta(Q2, Pr)=Q1$	$\delta(Q2, Pp)=Q1$
	$\delta(Q2, Ps)=Q2$	$\delta(Q2, Pr)=Q2$	$\delta(Q2, Pp)=Q2$
	$\delta(Q2, Ps)=Qe$	$\delta(Q2, Pr)=Qe$	$\delta(Q2, Pp)=Qe$
Q1	$\delta(Q1, Ps)=Q1$	$\delta(Q1, Pr)=Q1$	$\delta(Q1, Pp)=Q1$
	$\delta(Q1, Ps)=Qe$	$\delta(Q1, Pr)=Qe$	$\delta(Q1, Pp)=Qe$
Qe 終了状態			

実際に状態遷移表と遷移関数を利用する時に、前述したRule-1～Rule-6に従うほか、文字溢れ状態を確定する為に、前記定義3によって文字の溢れ率(Rf)を再計算しなければならない。再計算された文字の溢れ率(Rf)により、表1から検索し、文字溢れ状態を確定する。次は、新しく確定された文字溢れ状態により、次の調整動作(決定性有限オートマトンのΣ有限のアルファベット、即ち入力記号の有限集合に相当する)を決める。調整は完了するまで繰り返して行なう。

3.3.3. 調整アルゴリズム

前節で定義した状態遷移表と調整ルールにより、多国語対応自動組版の文字調整のアルゴリズムが纏められる。

- ① Qs (初期状態) = Qe の場合、処理を終了する。
- ② Qs (初期状態) において、Rf0 を算出する。
- ③ Rf0 (初期状態の溢れ率) と Rule-3～Rule-5 により、P1 (Ps、Pr、Pp の何れか) を選べる。P1 を確定できない場合、Qe 状態に遷移し、今まで調整を行った序列 (Pi、i=1) は受理不能序列とみなす。受理不可序列を記録する。
- ④ Pi (Pi、i=1, 2, ...) の処理 (文字調整) を行ない、Qi (Q1、Q2、Q3、Q4、Q5、Qe の何れか) を得る。
- ⑤ Qi が Q1、Q2、Q3、Q4、Q5 の遷移状態の場合、新たに算出された Rfi (溢れ率) と Rule-3～Rule-5 により、次の Pi を選べる。Qi の状態で Pi を確定できない場合、Qe 状態に遷移し、今まで調整を行った序列 (Pi、i=1, 2, ...) を受理不能序列とみなす。Pi を確定できる場合、処理を繰り返す。
- ⑥ Qi = Qe の場合、処理を終了する。今まで調整を行なった序列 (Pi、i=1, 2, ...) を受理可能序列とみなす。必要な場合、受理可能序列を記録する。

3.4. 調整不能の処理

調整不能の場合、受理不可序列を分析し、調整ルール及び容認された調整限度 (Rsm、Rrm、Rpm) を変更してから、再度調整を行う。

4. 逆調整処理

本論は訳文が長くなる場合の収納処理を中心に論じたが、訳文が短くなる場合、印刷のフォーマットが変わらないように調整方向を逆に行なうこととなる。この場合、前述した諸項目の変動率 (文字サイズ、文字比率、文字詰め) の計算を逆にすれば (例えば：加算→減算、減算→加算)、文字溢れないように調整すれば、訳文が短くなる場合でも、オートマトン方式を利用できるのである。

5. 結論

多国語対応の自動組版の文字調整を全て倍率単位で行なっているので、テキストフレームの溢れ率に基づいて、オートマトンのモデルを抽出することができた。更にオートマトンのアルゴリズムにより適当な強さを持つ調整項目を選べるので、より知能のある調整方式が実現された。

記録されたオートマトンの受理可能序列を取得して、同じようなテンプレートを調整する時に、自由調整序列方式の入力序列として利用するのに役に立つ。

6. 謝辞

本論文は2005年3月に初稿を完成したが、この度、電子情報通信学会・デジタルドキュメント研究会で発表する為に、本論文を大幅に訂正する際に、当社の堀事業部長、河村室長、上田部長等のご指導とご応援を頂き、深く感謝する。

文 献

- [1] 米田政明・広瀬貞樹等著「オートマトン・言語理論の基礎」 近代科学社
- [2] A. サローマ著「計算論とオートマトン理論」野崎昭弘等訳 サイエンス社
- [3] 守屋悦朗著 「形式言語とオートマトン」(Information Science & Engineering-F 3) サイエンス社
- [4] アドビ社 InDesign SDK 技術文献「INDESIGN® CS/INCOPIY® CS PROGRAMMING GUIDE」
- [5] アドビ社 InDesign SDK 技術文献 「Making Your 1st InDesign Plug-in」