

文書の文字内容割付けに関する一考察

4J-4

高杉正勝* 藤川佳三* 村上晴夫**

*日立西部ソフトウェア(株)

** (株)日立製作所 ソフトウェア工場

1. はじめに

文書は次の2つの体系に分ける事ができる。(OSIのODAではこの体系)

(1) 文書体系

文書の構成体である章や節、ページやページ内の区画等の特性(大きさや配置に関する属性)およびそれらの関係に関する規則である。この規則に従って区画の配置やページへの割付け内容の決定が行われる。

(2) 内容体系

内容要素である文字や画素等の特性(表現形式や内容の配置に関する属性)やそれらの関係に関する規則である。この規則に従って区画内に内容要素が配置される。

内容要素および制御機能は、内容体系によって定められる内容部を構成する。どの内容部も、文書体系によって定められる文書を構成する基本的な部品(割付けの面から見ると区画)と結び付けられる。

本稿では、このように体系付けた場合の文字内容割付けに関する問題点を上記の体系をもつOSIのODAの規格(ISO8613)を例に、考察する。

なお、ISO8613では各体系に応じた処理のモデルを定めており、それぞれを文書割付け処理、内容割付け処理と呼んでいる。

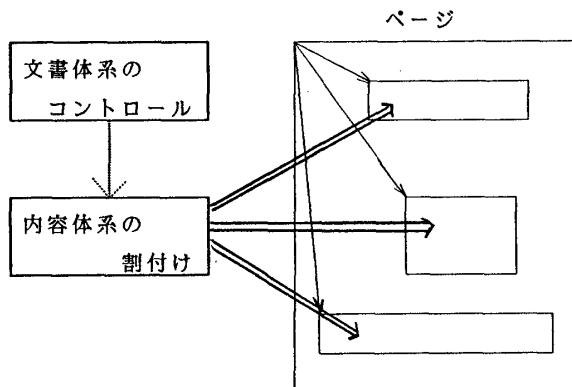


図1 文書体系・内容体系の関係

2. 文字内容割付けに関する問題点

文字内容の割付けに関しては、上記のように体系づけた場合、次のような割付け時に問題となる。

(1) 行の等間隔割付け

(2) マルチカラム時の行の同期割付け

以降、それぞれの割付けについて説明しこれらの問題点について述べる。

2. 1行の等間隔割付け

一般的に、日本語文書の場合、ルビや添え字、肩文字が行に含まれていても、その行と前後の行の間隔は一定である。このように割付けることを「行の等間隔割付け」という。(行間隔とは、参照線間の距離)

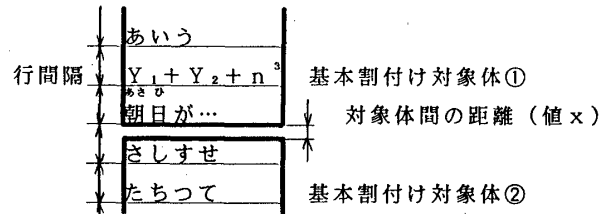


図2 行の等間隔割付け

2. 2マルチカラム時の行の同期割付け

行の同期とは、図3に示すように行を複数のカラムで行(参照線)を描いて割付けることをいう。

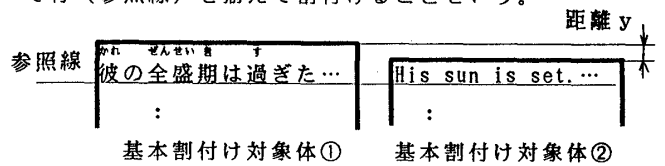


図3 行の同期割付け

2. 3 問題点

これらの、割付けの実現には、次の二つの情報が関係している。

・基本割付け対象体間の距離

割付け属性「分離」をもとに文書割付け処理が実際の基本割付け対象体間の距離を決定する。

・表示属性「初期オフセット」

基本割付け対象体の可視化の際の初期点の位置
内容割付け処理が先頭行の内容および他の表示属性との関係から決定する。

これらは、規格(ISO8613-2,6)をみると、それぞれの情報は独立した体系となっているが、文字割付けに関しては、実に密接に関係している。

例えば、図2のような場合、基本割付け対象体間の距離は、 x だけとる必要があるが対象体②の先頭行の内容がルビを含んでいるような場合、距離は0でよい。(ただし、ルビ文字のフォントサイズ+通常文字のフォントサイズ=行間隔の場合)

また、対象体①の最下行が添え字を含んでいるような場合も、対象体間の距離が変わってくる。

すなわち、対象体間の距離はその境界の行の内容により決定され、しかもその行の内容を割付けてみなければわからない。

よって、内容を割付けるための割付け可能領域のサイズを文書割付け処理で決定できず、そのために次の対象体に入るであろう内容を1行割付けてみる必要がある。

図3の場合、基本割付け対象体②の内容を割付けても、文書割付け処理では対象体①の内容が判らないため参照線を揃えるための距離 y が決定できない。

また、距離 y が決定できないと、内容を割付けるための割付け可能領域のサイズが決定できず、図2と同様の問題が発生する。

3. 考察

図4に示すように主たるフォントサイズが変わらない場合でも、その内容(ルビ、肩文字、添え字を含むか否か)によって $s1$ から $s6$ のように行枠の大きさが変化する。

すなわち、基本割付け対象体の大きさもその先頭行および最下行の内容によって変化する。(対象体の大きさ<行進行方向> \neq 行間隔 \times 行数)

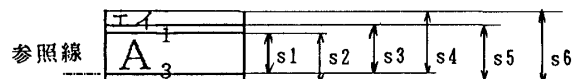


図4 内容による行枠の大きさの変化

これら行枠の情報(大きさや行枠前方範囲、行枠後方範囲)を決定するのは内容割付け処理であり、それら割付け情報は内容割付け処理で閉じた情報となっている。

しかし、ページの中に適切な行配置を行うには、文書割付け処理と内容割付け処理間で行枠情報の受け渡しが必要である。(少なくとも、先頭行と最下行の情報)

すなわち、内容割付け処理で、これから割付ける基本割付け対象体が、どのような行のつぎにくるか意識するか、あるいは文書割付け処理で行枠の情報を意識するかのいずれかの意識が必要である。このように互いの情報が相互に作用しあって値が決まるものがある。

これらは前述のように分けた場合の体系づけの問題点であり、新たな体系付けが必要であると考えられる。

現在の文書・内容体系にさらに、文字内容については文書体系で行枠の概念を加えた考え方が必要であると考えられる。

4. 参考文献

- 1) ISO8613-1~8, Information processing - Text and Office Systems, Office Document Architecture (ODA) and Intechange Format(1988).
- 2) 棟上昭男他: O S I の応用, pp. 84-111, 日本規格協会(1987).
- 3) J. Andre他: STRUCTURED DOCUMENTS, CAMBRIDGE UNIVERSITY PRESS(1989).